Pure Appl. Geophys. 176 (2019), 1037–1059
© 2018 Springer Nature Switzerland AG
https://doi.org/10.1007/s00024-018-1995-6

Pure and Applied Geophysics

Earthquake Fingerprints: Extracting Waveform Features for Similarity-Based Earthquake Detection

Karianne J. Bergen[1,2] and Gregory C. Beroza[3]

Abstract—Seismologists are increasingly adopting data mining and machine learning techniques to detect weak earthquake signals in large seismic data sets. The detection performance of these new methods, especially their sensitivity and false detection rate, depends on the choice of feature representation for waveform data. We have previously introduced Fingerprint and Similarity Thresholding (FAST), a new method for waveform-similarity-based earthquake detection that uses a pattern mining approach to detect earthquake signals without template waveforms. FAST has two key steps: fingerprint extraction and efficient indexing for similarity search. In this work, we focus on FAST fingerprint extraction: the method used to map short-duration waveforms to a set of features, called waveform fingerprints, used for detection. We describe the FAST fingerprint extraction method, a data-adaptive variation on the Waveprint audio fingerprinting method tailored for use in continuous seismic data. We compare the performance of the FAST fingerprint extraction method with existing fingerprinting techniques designed for audio identification. To overcome the challenges associated with using limited or incomplete event catalogs to evaluate detection algorithms, we propose a framework for quantifying the performance of different fingerprint extraction methods in the context of blind similarity-based detection. Our framework uses computational experiments on benchmark data sets, constructed with known event waveforms, to compute a measure of fingerprint effectiveness. We use this framework to show that, among the audio fingerprinting schemes considered in this work, our proposed FAST fingerprint extraction method achieves the most consistent performance in distinguishing similar, low signal-to-noise earthquake waveforms from noise in waveform data sets from eight stations in the Northern California Seismic Network.

Key words: Earthquake detection, seismology, audio fingerprinting, feature extraction, time-series analysis, similarity search.

1. Introduction

Earthquake detection, the task of identifying earthquake signals in continuously recorded ground motion data from one or more stations in a seismic network, is a challenging and fundamental task in seismology. Waveform cross-correlation, also referred to as template matching, is a widely used and highly sensitive method for detecting weak earthquake signals (Gibbons and Ringdal 2006). Waveform cross-correlation is limited to detecting events with waveforms similar to those of known events, and these template waveforms are derived from earthquake catalogs that are often incomplete. To overcome this limitation, Yoon et al. (2015) introduced Fingerprint and Similarity Thresholding (FAST), a general detector based on waveform similarity. FAST[1] is the first detector based on waveform similarity that is capable of identifying events with unknown sources in long-duration (large-T) data sets of up to 10 years (Rong et al. 2018).

FAST is inspired by algorithms used for content-based audio search and retrieval tasks, such as identifying songs directly from noisy audio waveform data recorded on a mobile phone (Wang 2003; Baluja and Covell 2008). FAST is an uninformed (unsupervised) detector and does not require template waveforms. The FAST detection model, like waveform cross-correlation, is based on the observation that earthquakes with similar sources produce similar waveforms. In the absence of template waveforms, FAST labels any pair of waveforms with high similarity as candidate earthquakes, relying on the assumption that noise waveforms are mutually

[1] Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA. E-mail: karianne_bergen@fas.harvard.edu

[2] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA.

[3] Department of Geophysics, Stanford University, Stanford, CA 94305, USA. E-mail: beroza@stanford.edu

[1] FAST code available at https://github.com/stanford-futuredata/FAST.

dissimilar (in contrast, waveform cross-correlation requires high similarity to a *known* template waveform). FAST identifies candidate earthquakes by using locality-sensitive hashing (LSH) (Andoni and Indyk 2006), a randomized algorithm for computationally efficient similarity search, to find pairs of similar waveforms in single-channel continuous data.

The FAST detection pipeline is made up of algorithmic modules, each of which can be independently tuned or optimized. Single-station FAST has two key steps: fingerprint extraction and efficient similarity search (see Fig. 1); an overview of the single-station FAST detection method is presented in Yoon et al. (2015). In the fingerprint extraction step, FAST computes a set of features, called a *waveform fingerprint*, for each short-duration sliding window in the single-channel continuous waveform data, and performs similarity search over this collection of waveform fingerprints. Waveform fingerprints are used as proxies for the raw waveform signals in the FAST similarity search step. Therefore, the performance of FAST will depend on our ability to extract appropriate features as inputs to similarity search for unsupervised earthquake detection. This work focuses on the fingerprint extraction step and extends the preliminary results of Bergen et al. (2016).

The efficient similarity search step, which is optimized for large-T detection in Rong et al. (2018), depends only on the choice of similarity metric for the waveform fingerprints, and not on the feature extraction algorithm itself. Bergen and Beroza (2018) present a multi-station extension to FAST, which combines the single-station FAST detection results to reduce false detections among the candidate events. Extending FAST over a network requires running FAST separately on data from each station, so the FAST feature extraction method is designed for use on single-station data [multi-station feature extraction is addressed in the supplement of Bergen and Beroza (2018)].

Earthquake detection is an active area of research, and in recent years there has been growing interest in machine learning and data mining approaches for event detection and seismic signal analysis. The performance of these methods can be critically dependent on the selection of appropriate waveform features. In the related field of audio signal processing, there is an extensive body of literature dedicated to methods for extracting features, called audio fingerprints, from raw audio waveform data for audio classification and identification (Cano et al. 2005). However, the topic of feature extraction for seismic

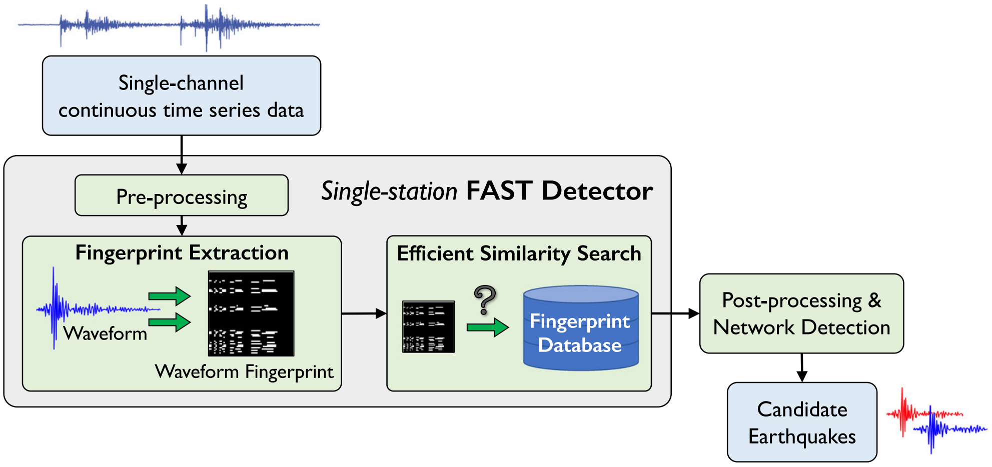

Single-channel continuous time series data
Single-station FAST Detector
Pre-processing
Fingerprint Extraction
Waveform
Waveform Fingerprint
Efficient Similarity Search
?
Fingerprint Database
Post-processing & Network Detection
Candidate Earthquakes

Figure 1
Main processing steps in the FAST earthquake detection method (Yoon et al. 2015). The input to FAST is continuous ground motion data from a single channel, usually with some basic pre-processing such as bandpass filtering. The FAST detector has two key steps: Fingerprint Extraction, which is the subject of this work, and Efficient Similarity Search. The Efficient Similarity Search step uses locality-sensitive hashing (Andoni and Indyk 2006) to build an index or "database" of waveform fingerprints and query the database to identify similar waveforms (Yoon et al. 2015; Bergen et al. 2016; Bergen and Beroza 2018; Rong et al. 2018)

signals has received less attention. Valentine and Trampert (2012) analyze the appropriateness of autoencoder neural networks for extracting compact waveform features from which the earthquake waveforms can be reconstructed. Holtzman et al. (2018) use an unsupervised learning approach to extract fingerprints for a set of earthquake waveforms that reveal subtle differences and relationships between events. Both studies apply their analysis only to event waveforms rather than to continuous data, and neither focuses on the task of discriminating earthquake signals from noise.

The focus of this work is feature extraction for the specific task of earthquake detection via blind similarity search in long-duration continuous seismic waveform data sets. We discuss the desirable properties of waveform fingerprints in the context of similarity search, and present the fingerprint extraction method used in FAST and alternate methods from the audio fingerprinting literature. One of the inherent challenges in developing new earthquake detection algorithms is the lack of ground truth data, an objective reference data set for measuring algorithm performance. Existing event catalogs, the closest thing to such a reference data set that is available, are known to be incomplete and thus do not accurately label all segments in the continuous waveform data as either earthquake signals or noise/non-earthquake signals. Without an objective performance metric, it is difficult to assess and compare the performance of different detection algorithms. Thus, in order to evaluate the relative performance of different fingerprinting schemes, we introduce a framework for quantifying the relative effectiveness of different feature extraction methods for similarity-based detection using computational experiments. We use our proposed framework to demonstrate that the procedure used to extract FAST waveform fingerprints produces a lower false alarm rate and enables the detection of lower signal-to-noise events than the alternative methods considered. Our analysis framework provides a template for other researchers who wish to optimize and validate feature representations of seismic waveform data for use in machine learning or data mining techniques.

2. *Background*

2.1. *Fingerprinting*

Feature extraction refers to the process of mapping an input data object to a set of features. A *fingerprint* is a compact signature that represents an object, such as a data file or high-dimensional data vector (Broder 1993). *Fingerprinting* or *fingerprint extraction* is a form of feature extraction that follows two general properties: (1) if two fingerprints, $f(x_1)$ and $f(x_2)$, corresponding to data objects x_1 and x_2, are different from each other (i.e. not identical), then the original data objects x_1 and x_2 are also different, and (2) if x_1 and x_2 are different, then there should be a low probability that they map to the same fingerprint:

$$f(x_1) \neq f(x_2) \Rightarrow x_1 \neq x_2 \tag{1}$$

$$\mathbb{P}[f(x_1) = f(x_2)] \ll 1 \quad \text{if } x_1 \neq x_2. \tag{2}$$

These properties allow fingerprints to act as nearly unique identifiers, similar to the most commonly understood definition of (human) fingerprints, the ridge patterns on human fingers (Pankanti et al. 2002). Fingerprints are often used in conjunction with hashing in information search and retrieval tasks; applications include document fingerprints for detecting plagiarism and duplicate webpages (Manber 1994; Broder 1997), *acoustic* or *audio fingerprints* for identifying audio recordings directly from the waveforms rather than from metadata (Cano et al. 2005), and molecular fingerprints for finding similar molecular structures in chemical databases (Willett et al. 1998).

2.2. *Waveform Fingerprints for Similarity Search*

In this work, we are interested specifically in *waveform fingerprints*, features that represent short-duration seismic waveform data, and are used as inputs to the efficient similarity search step in FAST for waveform-similarity-based earthquake detection. Waveform fingerprints should be tailored for use in similarity search, and to the particular characteristics of seismic data and the earthquake detection problem.

- Waveform fingerprints should distinguish earthquake signals from noise. In earthquake detection via similarity search, earthquake signals are distinguished from noise by high similarity to one or more other earthquake waveforms. The fingerprinting scheme should map similar earthquake waveforms to similar fingerprints and map noise waveforms to fingerprints with low similarity to other noise fingerprints; we describe fingerprinting schemes that satisfy this property as *discriminative* for similarity search. Note that features that work well for supervised waveform classification, where features should map earthquakes and noise waveforms into different regions in the feature space, may not be suitable features for use in similarity-based detection.

- The fingerprinting scheme must be paired with an appropriate similarity measure. The similarity measure should be appropriate based on the characteristics of the fingerprints (e.g. real-valued vs. binary, sparse vs. dense) and suitable for use in an efficient search method [e.g. LSH or k-d tree (Bentley 1975)]; examples include the Jaccard, Hamming, and cosine similarities (Leskovec et al. 2014), and l_p norms (Datar et al. 2004).
- A good fingerprinting scheme needs to be robust to the types of distortions that are expected in the data set of interest. For earthquake detection, fingerprints should be robust to high levels of noise and differences in signal amplitude. In this work we focus on methods that extract features from the time-frequency representation (spectrogram) of the signal, as these are more robust to small phase shifts than time-domain features.
- The fingerprinting scheme should produce fingerprints that maintain similarity over long time periods, as activity along faults persists over long timescales, and ideally should be fixed for the full time duration. Care should be taken with adaptive or dynamic fingerprinting schemes to ensure that similar waveforms separated by long time periods will produce fingerprints with high similarity.
- The fingerprinting scheme should not require prior waveform information, since the FAST detection method is designed to be effective even when no template waveforms are available. The fingerprinting scheme may be either data-agnostic (i.e. not dependent on any properties or characteristics of the data) or data-driven, so long as it does not incorporate any label information.

3. Methods

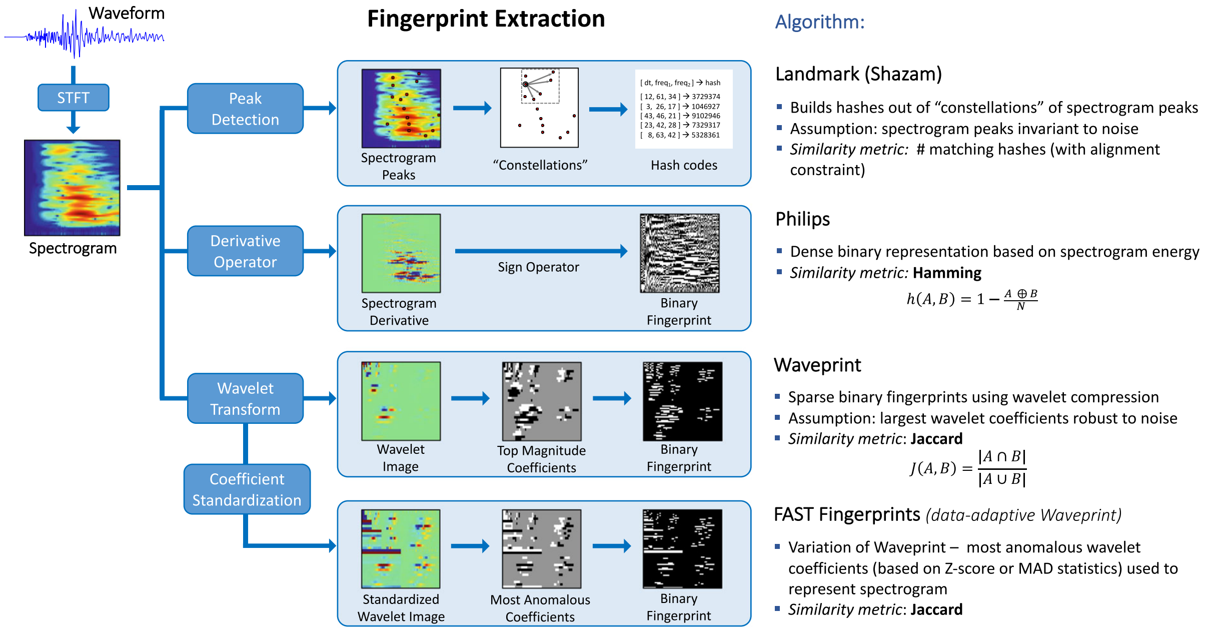
3.1. Audio Fingerprinting

Audio fingerprints are a compact representation of audio waveform data, including music, speech, or other sound recordings (Alías et al. 2016). Among previous work on fingerprinting for information retrieval and similarity search tasks, audio fingerprinting in particular provides a good starting point for the development of earthquake waveform fingerprints—there are structural similarities between the data (oscillatory signals with frequent zero-crossings and temporal variations in amplitude and frequency content), and both applications require fingerprints that are robust to small variations and additive noise.

In this work we focus on three well-known, effective audio fingerprinting methods: the landmark-based (Wang 2003), Philips (Haitsma and Kalker 2002), and Waveprint (Baluja and Covell 2008) methods. These methods and the FAST fingerprint extraction method are summarized in Fig. 2. In Sect. 5 we compare the performance of these fingerprinting schemes for seismic waveform data in the context of earthquake detection by blind similarity search.

3.1.1 Landmark Method

The landmark-based algorithm (Wang 2003, 2006), developed to identify audio clips recorded by a cell phone microphone for the Shazam mobile app, relies on the assumption that energy peaks in the time-frequency representation of the audio signal are key features that can be used to identify an audio clip and are invariant to additive noise. The algorithm identifies key points, defined as local maxima in the spectrogram, and builds hashes from the constellations based on the relative positions of these key points. The search and identification step scans a database to find sequences of multiple matching hashes.



Landmark (Shazam)

- Builds hashes out of "constellations" of spectrogram peaks
- Assumption: spectrogram peaks invariant to noise
- *Similarity metric:* # matching hashes (with alignment constraint)

Philips

- Dense binary representation based on spectrogram energy
- *Similarity metric:* **Hamming**

$$h(A,B) = 1 - \frac{A \oplus B}{N}$$

Waveprint

- Sparse binary fingerprints using wavelet compression
- Assumption: largest wavelet coefficients robust to noise
- *Similarity metric*: **Jaccard**

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

FAST Fingerprints *(data-adaptive Waveprint)*

- Variation of Waveprint – most anomalous wavelet coefficients (based on Z-score or MAD statistics) used to represent spectrogram
- *Similarity metric*: **Jaccard**

Figure 2

A comparison of three audio fingerprinting methods and the FAST feature extraction method. The landmark algorithm is similar to that used in the Shazam audio identification system (Wang 2003) and uses hash codes derived from peaks in the spectrogram. The Philips algorithm (Haitsma and Kalker 2002) is a dense binary fingerprint that uses the sign of the time and frequency derivatives to compute the fingerprint. The FAST feature extraction uses a modified version of the Waveprint (Baluja and Covell 2008) method

3.1.2 Philips Method

The Philips audio fingerprinting method (Haitsma and Kalker 2002) converts short segments of audio waveforms to a dense binary fingerprint. The Philips scheme computes the spectrogram, E, and then obtains a binary representation, F, from the sign of the derivatives of the signal energy; from Eq. (1) in Haitsma and Kalker (2002):

$$F[n,m] = \begin{cases} 1 & \text{if } (E[n,m] + E[n-1,m+1]) \\ & > (E[n,m+1] + E[n-1,m]) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This binary representation is then divided into fingerprints with a sliding window along the time axis. The similarity between these dense binary fingerprints (those with roughly equal numbers of 0s and 1s) is measured by the *Hamming similarity*. The Hamming similarity of bit strings (fingerprints) A and B, each of dimension N, is the fraction of the N bits that are identical in both strings:

$$\begin{aligned} \texttt{Hamming}(A,B) &= \frac{\#\{i:\ A[i] = B[i],\ i = 1,\ldots,N\}}{N} \\ &= 1 - \frac{A \oplus B}{N}. \end{aligned} \tag{4}$$

3.1.3 Waveprint Method

The Waveprint method (Baluja and Covell 2008) for audio identification takes an image processing approach to extracting audio fingerprints. Waveprint compresses sliding windows along the time dimension of the spectrogram (*spectral images*) using a discrete wavelet transform, a technique for image compression and denoising (Donoho and Johnstone 1994). The sign values of the largest wavelet coefficients are encoded in a sparse, binary fingerprint. This approach is robust against additive noise because the wavelet transform typically concentrates the signal energy in a relatively small number of wavelet coefficients, with the noise captured by *detail coefficients* that have smaller values and are not retained when the largest coefficients are selected.

3.2. Our Approach: FAST Waveform Fingerprints

The FAST waveform fingerprints are extracted using a modified version of the Waveprint feature extraction algorithm. The steps in the FAST feature extraction method and the modifications tailored for seismic data are discussed in the sections below.

In adapting audio processing techniques to seismic data, it is important to account for differences between the data, tasks, and constraints. Seismic event waveforms are relatively short in duration (typically seconds to tens of seconds), have a lower sampling rate (100 Hz vs. 44,100 Hz), and contain a narrower range of frequencies than audio data. Thus, overall, audio data tends to be richer and have more distinct spectral signatures compared to seismic data. Additionally, in the earthquake detection task, the signals of interest in the continuous waveform data tend to be infrequent, representing a rare class in a data set dominated by noise, while in audio identification tasks we do not expect a large number of fingerprints (either in the database or queries) to contain only noise.

3.2.1 FAST Fingerprint Extraction

The input to FAST is single-channel continuous waveform data. No template waveforms or label information is available to the algorithm. Each short-duration waveform is converted into a sparse binary fingerprint using the procedure described below (see Bergen 2018 for additional details).

0. *Data pre-processing* We apply a fixed bandpass filter to the continuous waveform data to remove frequencies that contain high levels of noise, especially narrow-band noise, or frequencies that do not contain useful information for the detection task.
1. *Spectrogram* We transform the time series data to the spectrogram, a time-frequency representation computed with the short-time Fourier transform. The spectrogram has a window length of w_s s and a lag between windows of length ℓ_s s. Although the original Waveprint algorithm uses logarithmically spaced frequency bins, for seismic data we use linear spacing in the frequency domain. We truncate the spectrogram frequency range when appropriate; inclusion of frequency bands containing high-amplitude noise or those outside the passband can hurt detection performance.
2. *Spectral images* We divide the spectrogram into short (w_f s) overlapping segments (ℓ_f s lag) and resize spectral images to fixed dimensions: d_1 frequency bins and d_2 time bins (e.g. $d_1 = 32$ and $d_2 = 64$). The dimensions should be fixed for all fingerprints in the data set, and each dimension should be a power of 2 (i.e. $d_1, d_2 \in \{x \mid x = 2^n$ for some integer $n\}$) to simplify the computation of the Haar transform in the next step.
3. *Haar wavelet transform* For each spectral image, we compute the two-dimensional discrete Haar wavelet transform (Mallat 2008) to produce a wavelet image. The Haar wavelet transform computes moving averages and moving differences at multiple scales in a signal or image.
4. *Coefficient standardization* We standardize the value of each wavelet coefficient, based on the distribution of values the coefficient takes over the full data set. For the ith fingerprint, the value of the jth wavelet coefficient, $X_j^{(i)}$, is replaced with a standardized value:

$$X_j^{(i)} \leftarrow \frac{X_j^{(i)} - a_j}{b_j}, \tag{5}$$

where a_j and b_j are measures of the center and spread, respectively, of the distribution of the jth wavelet coefficient. We consider two standardization schemes, one using the mean and standard deviation (Z score), and the other using the median and median absolute deviation (MAD) (Hampel 1974):

$$\begin{aligned} Z \text{ score:} \quad a_j &= \mu_j = \frac{1}{N_{fp}} \sum_{i=1}^{N_{fp}} X_j^{(i)}, \\ b_j &= \sigma_j = \sqrt{\frac{1}{N_{fp}} \sum_{i=1}^{N_{fp}} \left(X_j^{(i)} - \mu_j \right)^2}, \end{aligned} \tag{6}$$

$$\begin{aligned} \text{MAD:} \quad a_j &= \underset{i}{\mathtt{median}} \left(X_j^{(i)} \right), \\ b_j &= \underset{i}{\mathtt{median}} \left(\left| X_j^{(i)} - \underset{i}{\mathtt{median}} \left(X_j^{(i)} \right) \right| \right), \end{aligned} \tag{7}$$

where N_{fp} is the number of fingerprints in the data set. Before the statistics are computed, each of the wavelet images should be scaled to Frobenius norm 1. Because the properties of the noise-dominated continuous waveform data vary by station, it is necessary to perform the coefficient standardization using Z score/MAD statistics computed separately for each channel.

Coefficient standardization is a new step introduced for FAST feature extraction and is not included in the original Waveprint method. Based on the results presented in Sect. 5, the MAD statistics are used by default for coefficient standardization in FAST fingerprint extraction.

5. *Coefficient selection* We select the K *standardized* wavelet coefficients that are largest in magnitude; these correspond to the *most anomalous* Haar coefficients for each spectral image. K is typically selected in the range of 10–40% (which will produce a fingerprint sparsity of 5–20%).
6. *Conversion to binary* For the selected coefficients, we retain only the sign value and set all other coefficients to zero. We convert the sign values to binary using two bits per coefficient ($+ \rightarrow$ `01`, $- \rightarrow$ `10`, and $0 \rightarrow$ `00`), resulting in sparse binary fingerprints of dimension $D = 2 \cdot d_1 \cdot d_2$ with K non-zeros.

Because this fingerprinting scheme generates sparse, binary fingerprints, we measure the similarity between the resulting waveform fingerprints using a metric called the *Jaccard similarity*:

$$\texttt{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

Jaccard similarity is a similarity measure for comparing sets, here the sets of active (non-zero) coefficients in the fingerprints.

3.2.2 Haar Coefficient Standardization

The FAST fingerprint extraction method largely follows the Waveprint method, with the addition of the coefficient standardization step that is applied to the Haar wavelet coefficients prior to coefficient selection. We explored the use of other wavelet representations, but none of them exceeded the performance we obtained using the Haar basis. The necessity of the coefficient standardization step for the FAST fingerprints comes down to a key difference between seismic and audio data: in continuous seismic data, noise-only signals make up the majority of the data set (and, by extension, the majority of waveform fingerprints), while the signals of interest, earthquakes, are infrequent. FAST identifies candidate earthquakes by searching for all pairs of similar waveform fingerprints. Since most of the fingerprints extracted from continuous data correspond to noise signals, it is critical that noise fingerprints have low mutual similarity for FAST to be effective.

The original Waveprint method represents a waveform using the largest-magnitude wavelet coefficients, but when applied to noise-dominated seismic data sets this representation is inefficient; a small subset of wavelet coefficients regularly take large values and are frequently selected, while a larger number of coefficients are almost never selected (see Fig. 3 for examples of coefficient distributions). In a representative data set with 10% sparsity, 16% of the coefficients are active in at least 1 in 4 fingerprints, while half of all coefficients are active in fewer than 1 in 100 fingerprints (Bergen et al. 2016). This inefficiency results in higher similarity between pairs of noise fingerprints (Fig. 4), and negatively impacts detection sensitivity by making it more difficult to identify pairs of similar low signal-to-noise (SNR) earthquake waveforms.

Coefficient standardization can be viewed as a balancing of the coefficient distributions, transforming the distribution of each coefficient to be centered around zero and have roughly the same spread. This increases the probability of selection for coefficients that originally were centered at zero with a small variance (or MAD), but decreases the probability of selection for coefficients that were originally centered away from zero or had large variance. The updated FAST fingerprint representation is more efficient: after MAD standardization is applied to the example data set described above, each coefficient is active in 5–15% of all waveform fingerprints, and there are no longer coefficients that are frequently or rarely active. The underlying assumption when we perform this balancing of the coefficient distributions is that it will not hurt detection performance by substantially reducing the similarity between the fingerprints of similar earthquake waveforms; we show in Sect. 5

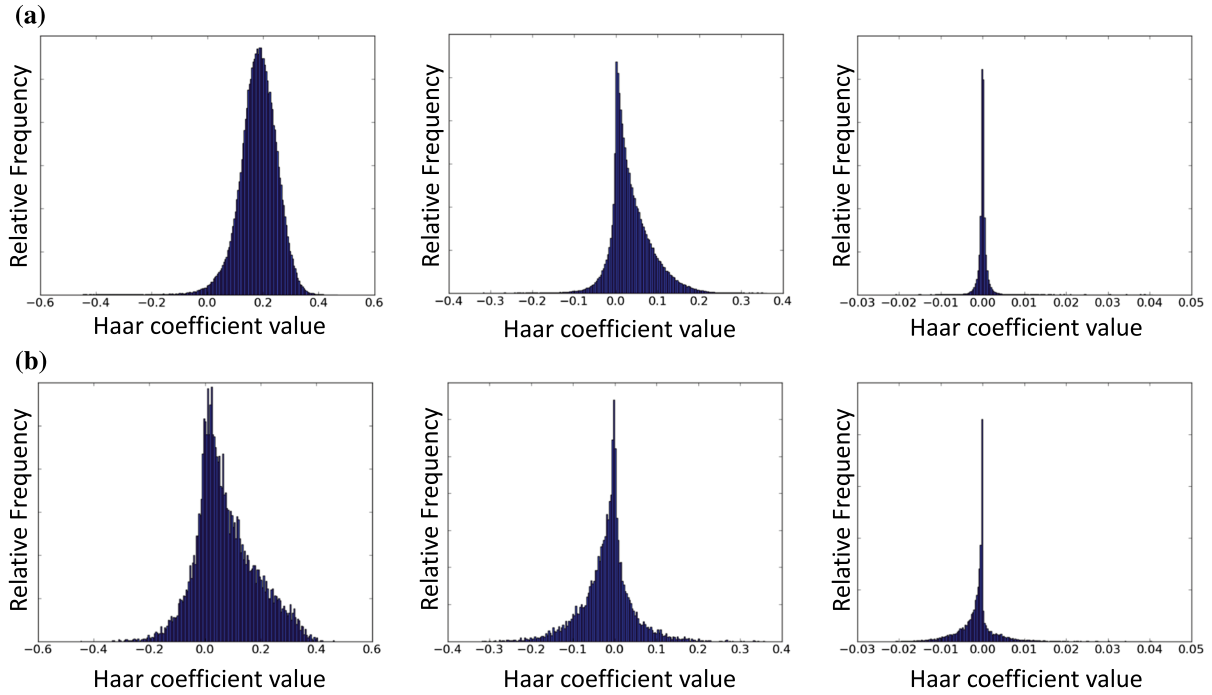


Figure 3

Distribution of wavelet coefficient values across all **a** noise and **b** earthquake fingerprints extracted for station NC.CCOB for 3 of the 2048 wavelet coefficients. Note that the same 3 wavelet coefficients (left to right: coefficient nos. 64, 322, and 273) are shown for both noise (top) and earthquake (bottom) waveforms; noise and earthquake waveform data were separated for comparison. For some coefficients there is not a significant difference between the coefficient distributions for noise and earthquake waveforms, but often the distribution for earthquakes has a different spread (right panel) or a different center or skew (center and left panels). For the purposes of coefficient standardization, each of these distributions is modeled by two parameters: the mean and standard deviations (*Z* score) or median and median absolute deviation (MAD). **a** Examples of distributions of wavelet coefficient values for noise waveforms. Most of the distributions are strongly peaked (right), but some are less strongly peaked (left) or asymmetrical (center). **b** Examples of distributions of wavelet coefficient values for earthquake waveforms. As in the case of noise waveforms, most of the wavelet coefficient distributions for earthquake waveforms are strongly peaked. The wavelet coefficient distributions for earthquakes often differ from the corresponding distribution for noise

that, empirically, this approach does preserve similarity for earthquakes (Fig. 4).

4. Experiments

In order to select the best feature extraction method and optimize any necessary parameter values (e.g. the sparsity parameter, *K*, for Waveprint and FAST fingerprints), we require a means of quantifying the performance of each method in the context of similarity search. In this section we describe a test for comparing the performance of different feature extraction methods. In the next section we use this benchmark test to compare three audio fingerprinting methods (Sect. 3.1) and the FAST feature extraction method for use in waveform similarity search.

4.1. Benchmark Data Set

To compare the effectiveness of different fingerprinting schemes for waveform similarity search, we need examples of known earthquake and noise waveforms for performance evaluation. The best fingerprint extraction method and optimal parameter settings can vary by station depending on the local noise properties, so we use multiple data sets to reduce bias and determine the method with the best overall performance. We compile a benchmark data set containing earthquake waveforms and segments of background noise from each of eight different

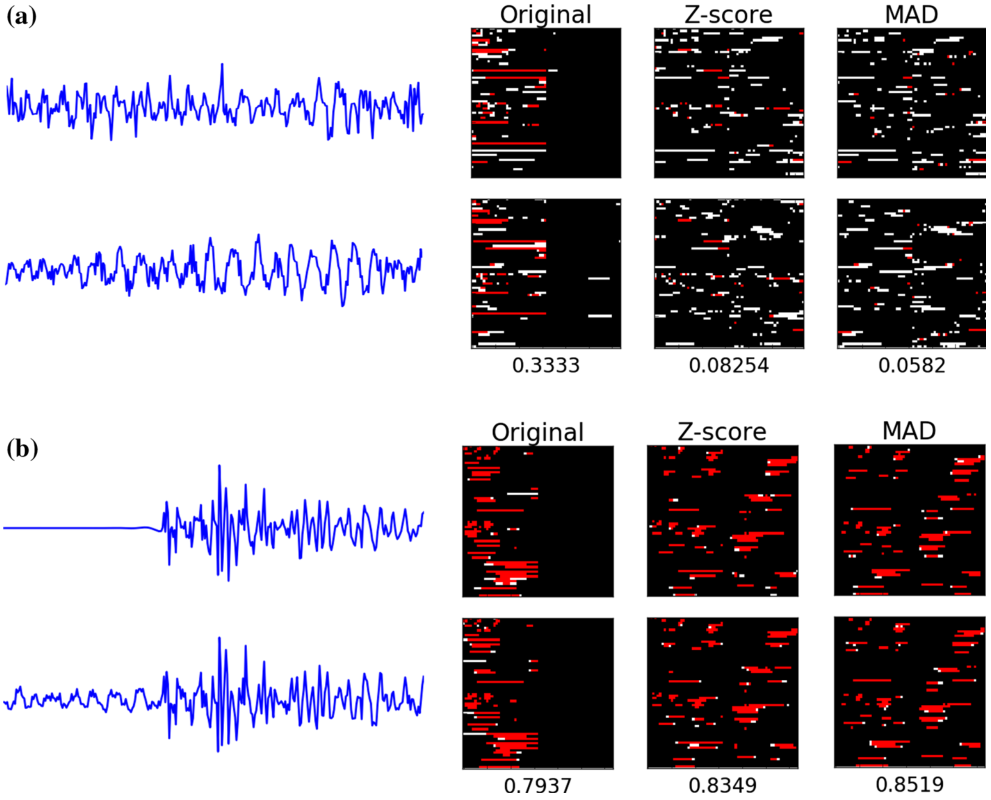


Figure 4

Comparison of Waveprint and FAST fingerprinting schemes applied to **a** background noise and **b** similar earthquake waveforms. Each row displays the fingerprints extracted from corresponding waveform data, and each column represents a different fingerprinting scheme: Original refers to the Waveprint scheme with no coefficient standardization step; *Z* score and MAD refer to standardization with the mean and standard deviation or median and median absolute deviation, respectively. Pixels in white and red represent the active/selected coefficients, with those in red indicating the intersection of fingerprints for each of the two distinct waveforms under the same fingerprinting scheme. All waveform data below are from station BK.SAO and represent a duration of 16 s. **a** Fingerprints corresponding to two noise signals. The Jaccard similarities between fingerprints are: 0.33 (original), 0.08 (*Z* score), and 0.06 (MAD). Fingerprints corresponding to noise waveforms have lower similarity after coefficient standardization compared to the original fingerprints. **b** Fingerprints corresponding to two similar earthquake signals: the first waveform corresponds to a real event; the second waveform is a copy of the event waveform embedded in noise at SNR 5.0 (as illustrated in Fig. 6). The Jaccard similarities between fingerprints are: 0.79 (original), 0.83 (*Z* score), and 0.85 (MAD). Fingerprints corresponding to similar waveforms maintain high similarity after coefficient standardization

stations. The stations, shown in Fig. 5, are selected for geographic diversity within the northern California region and to ensure availability of a relatively high number of event waveforms.

For each of the eight stations, we create an "earthquake" and a "noise" waveform data set. All data are taken from the 8.5-year period from 1 January 2008 to 31 May 2016. For both earthquake and noise waveform data, we use the vertical channel data, apply a 1–10 Hz bandpass filter, and decimate to 20 Hz. We apply the same bandpass filter for each station because we want to ensure good performance even if the optimal bandpass filter is not selected.

The earthquake waveform data is taken from catalog events in the Northern California Seismic Network (NCSN) phase-pick catalog (NCEDC 2014). We include only events magnitude 2.0 and larger. For each event in the NCSN catalog, we download a 2-min segment of waveform data (when available): the 60 s leading up to the P-wave arrival and the 60 s after the P arrival. To ensure that the earthquake data set contains events with clear waveforms, we include events only if (1) the wave arrival appears close to the

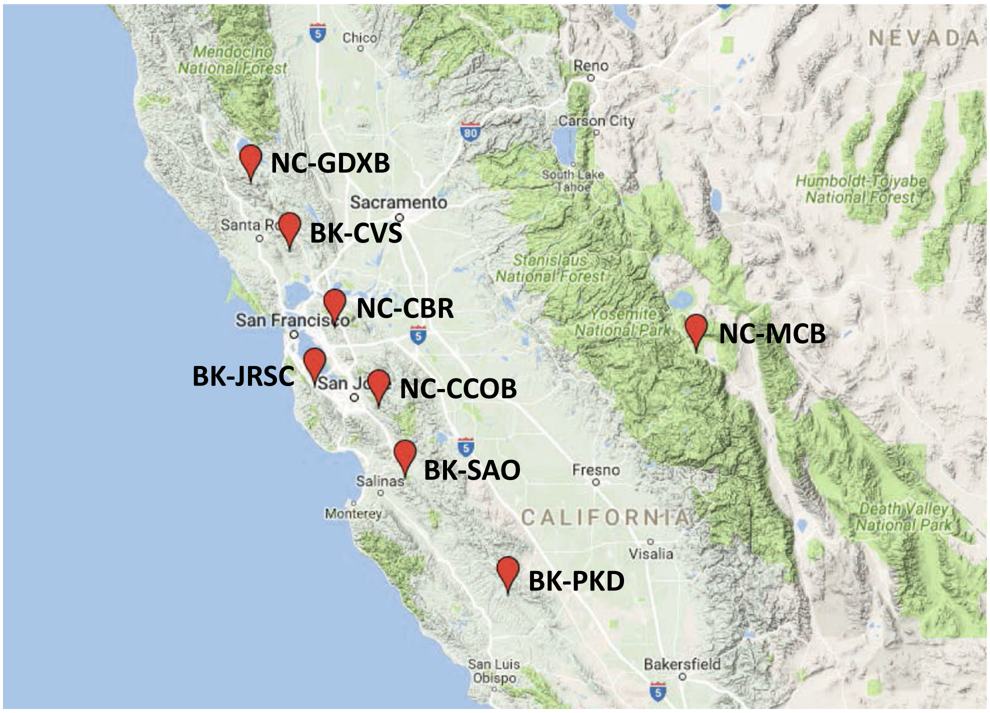


Figure 5
Locations of the eight selected stations used in benchmark data sets to compare performance of feature extraction methods: Parkfield (BK.PKD), Hollister (BK.SAO), Calaveras Fault (NC.CCOB), San Francisco Peninsula (BK.JRSC), San Ramon (NC.CBR), Napa Valley (BK.CVS), Mammoth (NC.MCB), and Geysers geothermal field (NC.GDXB). Created with Google Maps

Table 1
Benchmark waveform data sets

Station	Number of waveforms in data set	
	Earthquakes	Noise ($\mathcal{N}_\mathcal{C}$ and $\mathcal{N}_\mathcal{L}$)
BK.PKD	1972	3083
BK.SAO	2271	3094
NC.CCOB	1691	2851
BK.JRSC	1824	3092
NC.CBR	1069	2469
BK.CVS	3012	2974
NC.MCB	1276	3051
NC.GDXB	1991	3080
Total	15,106	23,694

"Earthquakes" refers to the number of medium-to-high SNR catalog events in the data set for each station. "Noise" refers to the number of "clean" or "lively" background noise segments (with likely earthquakes excluded) in the data set for each station

expected P arrival time and (2) the STA/LTA ratio (Allen 1982) (with window lengths 1 and 30 s) exceeds a threshold value of 5.0 at some time up to 15 s after the P arrival. The number of earthquake waveforms included in the benchmark from each of the selected stations is given in Table 1, and the event epicenter locations are shown in Figs. 12 and 13 in the Appendix.

The noise waveform data are sampled from the continuous data. For each day in the 8.5-year period, we select a 2-min time interval at random and download the continuous waveform data (when available). These "noise" segments may contain regional or teleseismic earthquakes or other transient signals, so we apply an additional data cleaning step. We label each segment as either "clean noise" ($\mathcal{N}_\mathcal{C}$), "lively noise" ($\mathcal{N}_\mathcal{L}$), or "non-noise" ($\mathcal{N}_\mathcal{S}$) (see Appendix 7.2). Non-noise are segments of data that likely contain signals from regional earthquake arrivals, earthquake codas, or teleseismic waves; these segments are removed from the noise data set. Clean noise segments are those with relatively uniform signal energy across the interval, while lively noise segments have more variation in signal energy.

4.2. *Extracting Waveform Fingerprints*

All of the fingerprint extraction methods initially convert the continuous waveform data into the time-frequency domain. The computation and processing of the spectrogram is the same across all methods, with the exception of the choice of window function. All spectrograms use a window length $w_s = 10.0$ s and a window lag $\ell_s = 0.1$ s, and have 32 linearly spaced frequency bins (33 for Philips to produce 32 bins in the derivative).

In the Philips, Waveprint, and FAST fingerprints we use a fingerprint window $w_f = 10.0$ s. The Waveprint and FAST fingerprints have dimension $D = 4096$ and the Philips fingerprints have dimension $D = 3200$. For the Waveprint and FAST fingerprints we vary the value of sparsity parameter, K, from 50 to 800 for parameter testing, but we fix the value $K = 400$ when comparing different fingerprint extraction methods. In the benchmark data sets, earthquakes are oversampled, so we use only the noise waveforms to compute the Z score and MAD statistics for the FAST fingerprints.

The landmark-based method computes hashes rather than fingerprints. Key points are selected by identifying local maxima in the spectrogram using a maximum filter (with a minimum distance of 1.5 Hz along the frequency axis and 1 s along the time axis). The hash constellations are formed between each key point and other local maxima within a target region that includes any local maxima that both occurs 1–10 s after the key point and has a frequency coordinate within a 4 Hz range.

4.3. *Performance Metrics*

To evaluate the performance of different fingerprint extraction methods for earthquake detection, we require a metric to quantify the degree to which a fingerprinting scheme is discriminative for similarity search. A fingerprinting scheme is discriminative if (1) under the fingerprinting scheme, two similar earthquake waveforms are mapped to fingerprints that have high similarity, and (2) an arbitrary pair of noise waveforms are mapped to fingerprints that have low similarity. We quantify these criteria using two metrics: fingerprint accuracy and baseline similarity, respectively.

4.3.1 *Quantifying Accuracy and Baseline Similarity*

In the discussion that follows, let $x^{(i)} \in \mathbb{R}^M$ be the ith earthquake waveform, represented in the time domain with M samples, and let $n^{(j)} \in \mathbb{R}^M$ be the jth noise waveform. Let $\mathcal{F} : \mathbb{R}^M \to \{0,1\}^D$ represent the fingerprint extraction operation, and let α be a scaling factor that controls the signal-to-noise ratio (α is not a fixed value but varies to produce the desired signal-to-noise ratio; see Appendix 7.1).

Fingerprint accuracy is a measure of the quality of the fingerprints of earthquake waveforms for similarity-based detection under additive noise. In our benchmark test we consider the challenging detection task of identifying two similar event waveforms, both at low SNR, because we would like our fingerprinting scheme to be discriminative for weak earthquake signals. We compare the fingerprints of two versions of the same earthquake waveform, $x^{(i)}$, (high signal-to-noise, from "earthquake" waveform data set) embedded in two different noise segments, $n^{(j)}$ and $n^{(k)}$, both at low SNR (see Fig. 6):

$$accuracy(i,j,k) = \texttt{sim}\Big(\mathcal{F}(\alpha_j x^{(i)} + n^{(j)}),\, \mathcal{F}(\alpha_k x^{(i)} + n^{(k)})\Big), \tag{9}$$

where noise segments $n^{(j)}, n^{(k)} \in \mathcal{N}_C$ are only drawn from the "clean noise" data set from the same station as the earthquake waveform $x^{(i)}$. Here, $\texttt{sim}(\cdot,\cdot)$ represents the relevant similarity measure: number of matching hashes for landmark method, Hamming similarity for the Philips method, and Jaccard similarity for the Waveprint and FAST fingerprints.

Waveform fingerprints should be effective even when the earthquake waveforms are not perfectly aligned; this is necessary for good performance in continuous data when the lag between adjacent fingerprints may be large compared to the sampling rate. To make our tests more representative of the detection challenges in continuous data, we compare fingerprints for similar waveforms with a time offset

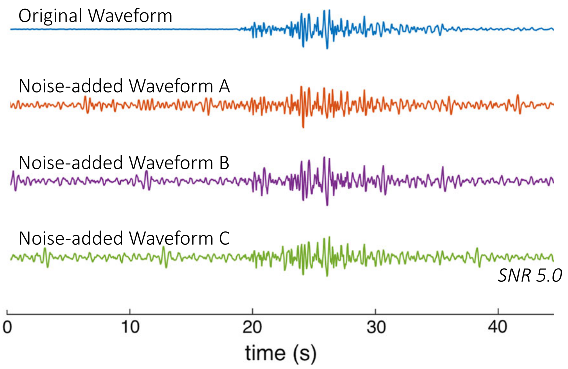


Figure 6
Original earthquake waveform is shown in the top row (blue). This original waveform is scaled down and embedded in three different intervals containing only sensor-recorded background noise, forming the three "noise-added" copies of the waveform in the bottom three rows, at a lower signal-to-noise ratio than the original waveform (SNR 5.0)

(applied before extracting fingerprints). We use offsets ranging from 0 to 10 samples, corresponding to the maximum offset for continuous data sampled at 20 Hz with a fingerprint lag of 1.0 s. These offsets only apply to the Philips, Waveprint, and FAST fingerprints, as the notion of a fingerprint lag is not applicable to the landmark-based hash method.

To measure the similarity between fingerprints extracted from noise waveforms, we define the *baseline similarity* between the kth and ℓth noise waveforms as:

$$\text{baseline}\ (k,\ell) = \mathtt{sim}\Big(\mathcal{F}(n^{(k)}),\ \mathcal{F}(n^{(\ell)})\Big), \qquad (10)$$

where noise segments $n^{(k)}, n^{(\ell)} \in (\mathcal{N}_C \cup \mathcal{N}_L)$ are drawn from both "clean " and "lively" noise. The baseline similarity is an important statistic because the threshold for what will be considered high accuracy for earthquake waveform fingerprints is determined relative to typical similarities between fingerprints for noise waveforms.

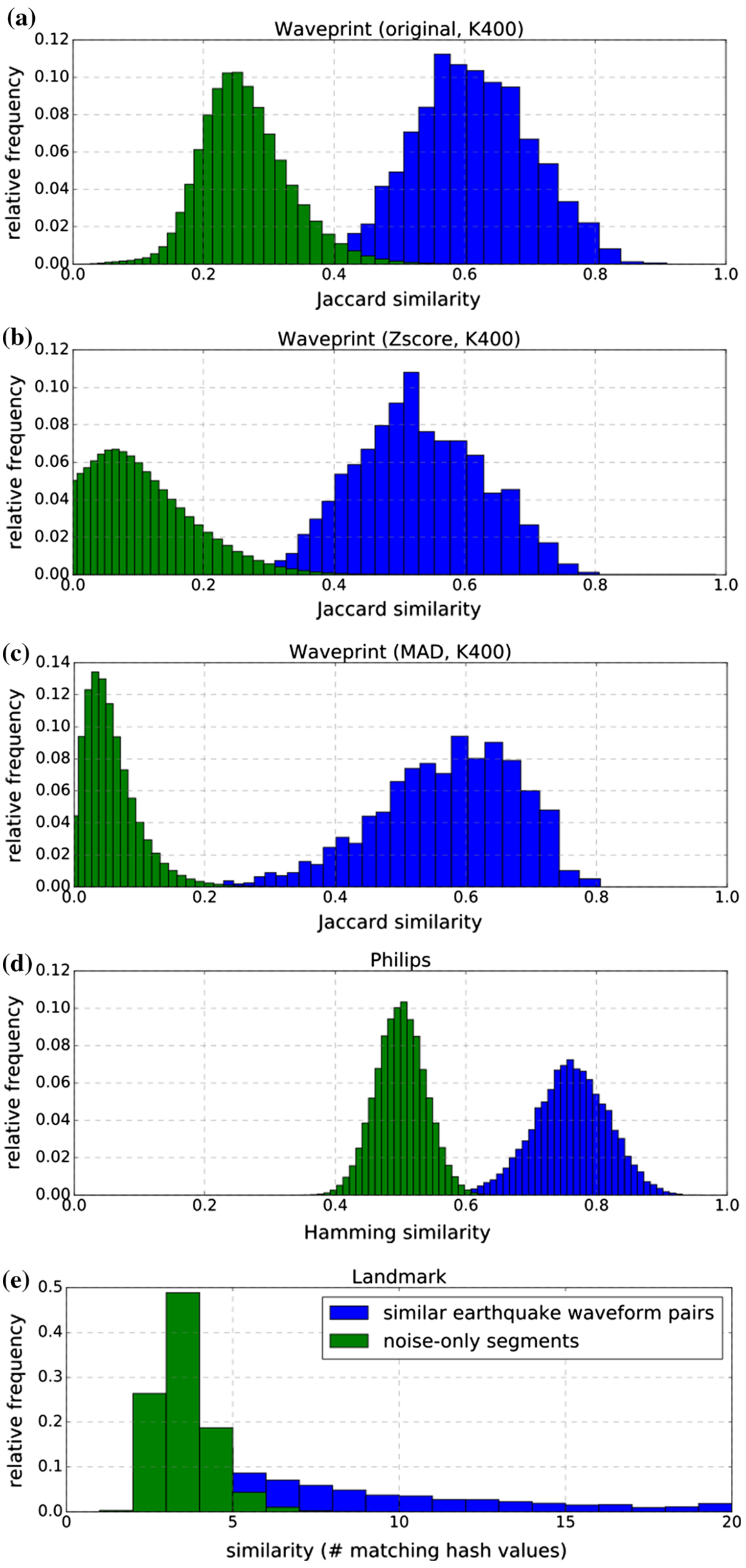
To obtain the baseline similarity distribution, we compare 1,000,000 pairs of noise fingerprints (some noise fingerprints may belong to multiple pairs) per station. For measuring the accuracy distribution, the number of pairs of fingerprints is proportional to the number of earthquake waveforms available from a given station (see Table 1). A separate distribution of values for fingerprint accuracy and baseline similarity are computed using waveform data from each of the eight stations.

Figure 7
Distributions used to characterize the trade-off between true positives and false positives. Green distributions show values of baseline similarity for pairs of noise fingerprints; blue distributions show accuracy measure for pairs of earthquakes embedded in noise at SNR 2.0. Data shown are from station BK.SAO. **a** Original Waveprint fingerprints (no coefficient standardization) with sparsity parameter $K = 400$. **b** FAST fingerprints (modified Waveprint with Z score coefficient standardization) with $K = 400$. **c** FAST fingerprints (modified Waveprint with MAD coefficient standardization) with $K = 400$. **d** Philips fingerprints. Because these fingerprints are dense binary fingerprints, the similarity of pairs of noise fingerprints is centered around 0.5, which is the expected Hamming similarity for two dense binary fingerprints with bits selected at random. **e** Landmark-based hashes. Similarity is quantified by the number of matching hash values, and is represented as counts rather than a quantity between 0 and 1 (as in the Jaccard similarity for Waveprint or Hamming similarity for Philips)

4.3.2 ROC Curve and Truncated-AUC

For a fingerprint extraction scheme to be effective for similarity-based detection in long-duration data, we require both high accuracy for fingerprints and low baseline similarity to limit false detections. We characterize the trade-off between false detections and missed detections using a receiver operating characteristic (ROC) curve. For a given Jaccard similarity threshold, τ, we define the *true positive rate* (TPR_τ) as the fraction of earthquake waveform pairs for which the accuracy exceeds the threshold, and we define the *false positive rate* (FPR_τ) as the fraction of noise waveform pairs for which the baseline similarity exceeds the same threshold (see Fig. 7). The true and false positive rates should be interpreted with caution with respect to detection in continuous data, as TPR and FPR do not account for the non-linear relationship between Jaccard similarity and the probability of detection in LSH-based similarity search. The ROC curve traces the values (FPR_τ, TPR_τ) for all choices of threshold τ between 0 and 1 for Philips and Waveprint variants (for landmark-based features, the threshold values are counts).

One commonly used measure to summarize the overall detection performance across the full range of threshold values is the area under the ROC curve (AUC) (Bradley 1997). Because high values for the FPR are not appropriate for the earthquake detection


(a)
Waveprint (original, K400)
relative frequency
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard similarity
(b)
Waveprint (Zscore, K400)
relative frequency
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard similarity
(c)
Waveprint (MAD, K400)
relative frequency
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard similarity
(d)
Philips
relative frequency
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.2
0.4
0.6
0.8
1.0
Hamming similarity
(e)
Landmark
relative frequency
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
similarity (# matching hash values)
similar earthquake waveform pairs
noise-only segments

task, we use a truncated version of the AUC that considers only the area under the section of the curve corresponding to small values of the FPR: FPR $\in [0, \mathrm{FPR}_{\max}]$; we normalize the truncated-AUC so that its maximum possible value is always 1 for any choice of $\mathrm{FPR}_{\max}$ (see Fig. 8).

5. Results

We compare the performance of audio fingerprinting and FAST fingerprinting schemes for earthquake detection: (1) landmark-based features, (2) Philips fingerprints, (3) Waveprint fingerprints, and FAST fingerprints with coefficient standardization using (4) *Z* score or (5) MAD statistics. Figure 9 shows a comparison of five fingerprinting schemes on the benchmark waveform data sets for eight stations using the truncated-AUC metric.

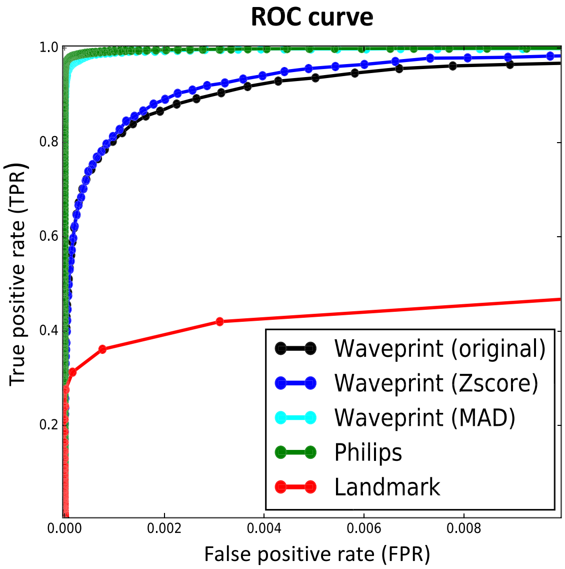


Figure 8
Comparison of ROC curve for different fingerprinting schemes. The area under each curve (with appropriate normalization) gives the truncated-AUC (FPR ≤ 0.01) statistic used in Fig. 9 for each method. Data shown are for the data set from station BK.SAO with earthquake waveforms at SNR 2.0. A separate ROC curve is computed for each station and method. ROC curves are generated from distributions in Fig. 7 by plotting the fraction of baseline similarity (green) distribution exceeding a given threshold to the fraction of the fingerprint accuracy (blue) distribution exceeding the same threshold, for all thresholds from 0 to 1 (or 0–20 for landmark method) for each method

At a higher SNR of 5.0, the Philips method and all Waveprint variants perform consistently well across the eight benchmark data sets. As the SNR is lowered to 2.0 and 1.0, the performance of all of the fingerprinting schemes degrades, and for some schemes there is more variation in performance, with a large performance gap between the best- and worst-case performance on the eight benchmark data sets.

At lower SNR (1.0–2.0), FAST fingerprints (MAD) achieve the best and most consistent performance; this scheme has the best worst-, average-, and best-case performance compared to the other methods. The Philips fingerprints also perform well in the low SNR case. At low SNR, the Waveprint and FAST fingerprints (*Z* score) achieve similar performance on average, but the FAST (*Z* score) scheme has more variation in performance compared to the original Waveprint method; on some data sets the FAST (*Z* score) scheme performs well, but on others its performance is very poor. In all cases, the landmark-based method performs poorly, even for signals in the higher SNR case. The box plots in Fig. 9 demonstrate that there can be significant variation in performance for fingerprinting schemes applied to data from different stations. The ranking of which fingerprinting schemes have the strongest performance may also vary by station.

Note that caution should be taken in interpreting the false and true positive rates for continuous data. The TPR and FPR are computed on the true Jaccard similarity values, and do not account for the fact that the probability of detection varies with the Jaccard similarity in FAST's LSH-based similarity search. Using the Jaccard similarity values gives a measure of the fingerprinting scheme effectiveness that is independent of the choice of the similarity search method or parameters, making it possible to draw general conclusions about the relative effectiveness of different approaches.

5.1. Parameter Testing

The approach that we have outlined above can also be used to compare and optimize parameter values for a given fingerprinting scheme. The performance of the FAST (MAD) fingerprinting scheme is not particularly sensitive to changes in

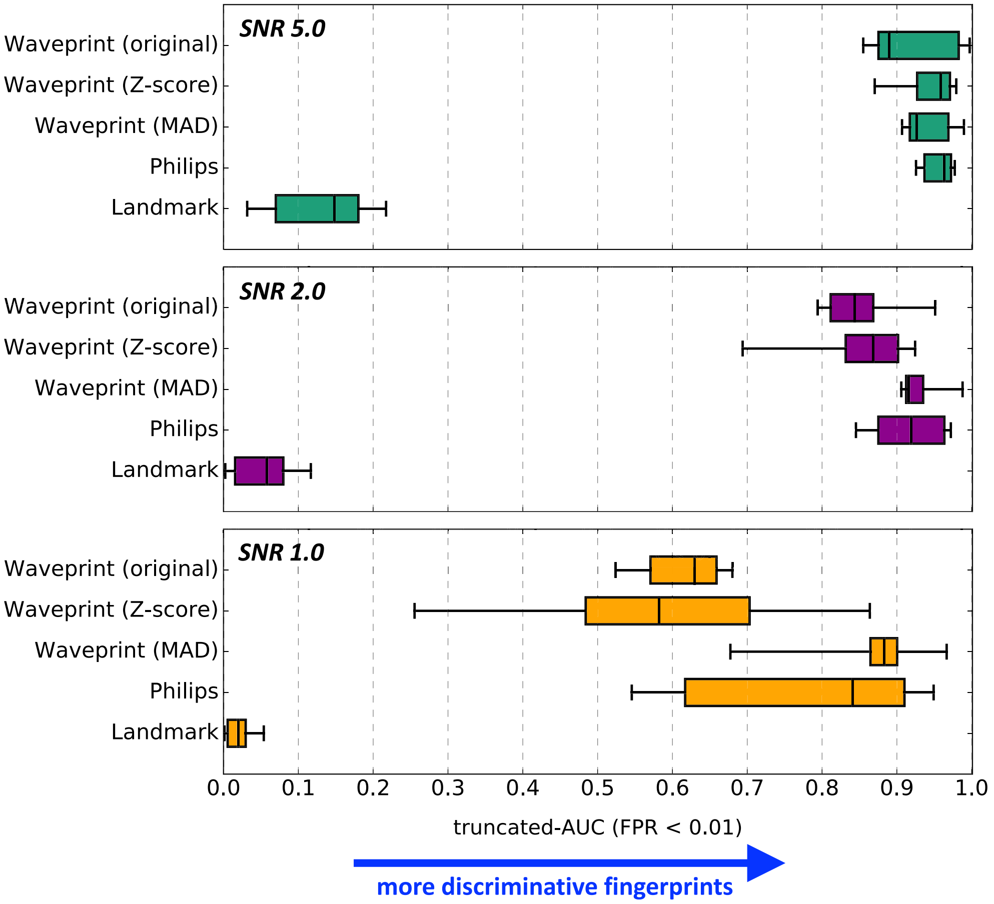


Figure 9
Relative performance of five fingerprinting schemes on the benchmark data sets. Box plots show variation in truncated-AUC measure over the data sets from eight stations. Each panel shows results for different signal-to-noise ratio (SNR) values for the earthquake signals: 5.0, 2.0, and 1.0 (top to bottom)

most of the parameter values (e.g. the choice of wavelet basis), with the exception of one key parameter: the sparsity parameter K. Ideally, fingerprints should be as sparse as possible (i.e. have few non-zero values, low K) to enable efficient search, but must remain sensitive to weak earthquake signals. A comparison of the performance with different values of K (Fig. 10) shows that values in the range of $K = 300$–500 (out of 4096 coefficients in binary fingerprint), corresponding to 10% sparsity, are optimal for the FAST fingerprints with MAD-adjusted coefficients.

6. Discussion and Conclusions

6.1. Audio Fingerprints for Earthquake Detection

Among the three audio fingerprinting methods, the landmark-based features perform poorly on seismic data even at high SNR, while both the Philips and Waveprint fingerprints perform well when applied to seismic data. Each of these three approaches captures spectrogram features at different scales; the landmark-based features only capture a few prominent points, the Philips fingerprints contain only local information (derivatives) in the spectrogram, and Waveprint captures information about different-sized patches (from local to global) within the spectrogram.

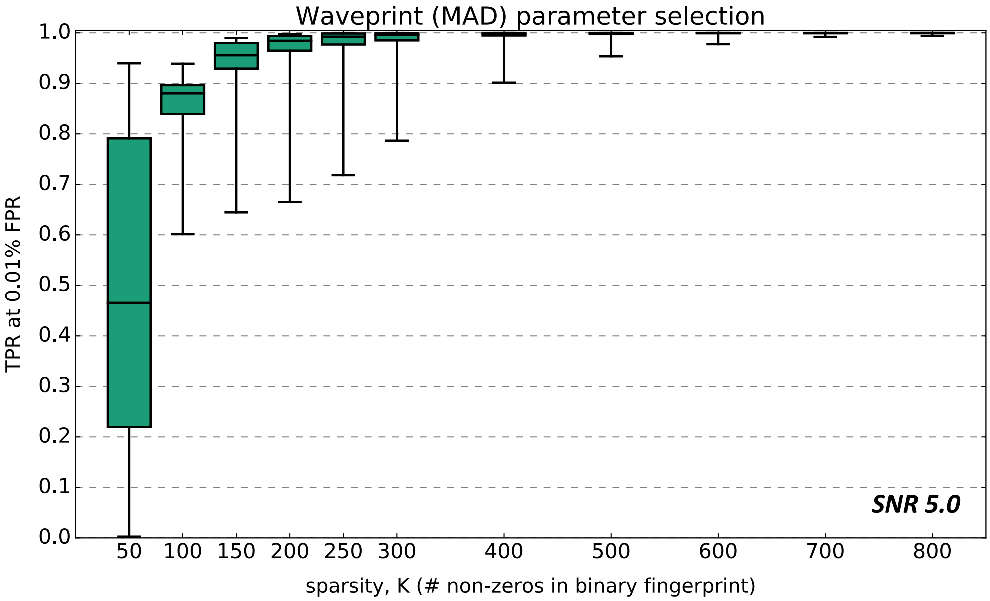


Figure 10

Parameter testing: the horizontal axis is the value of the sparsity parameter, *K*, which corresponds to the number of non-zero values in the binary fingerprint (out of 4096). As *K* increases, the fingerprints are less sparse, which makes them less compact to store, and the runtime of the efficient similarity search step of FAST tends to increase. The optimal range for *K* is 300–500, which corresponds to 7–12% sparsity. The vertical axis is a measure of how discriminative the resulting fingerprints are, as measured by the true positive rate (TPR) for a fixed false positive rate of 0.01%. Box plots show variation in performance over the waveform data sets from eight different stations

The landmark-based features perform poorly because local maxima in the spectrogram are not appropriate features for characterizing earthquake waveforms. Even for two earthquake waveforms that are nearly identical in the time domain, there may be limited correspondence between the spectrogram peaks for the two events (Fig. 11). This approach misses important information for detection, as P-wave arrivals correspond to intervals with increasing signal energy rather than energy maxima, and thus generate very few key points.

The Philips fingerprints are also essentially local features based on the derivatives of the spectrogram. An advantage of the Philips method is that each feature contains only information from two neighboring frequency bins, so relatively few features will be affected by the presence of persistent narrow-band noise, making this approach robust to a poor choice of bandpass filter. In contrast, because the Waveprint fingerprints and FAST variant effectively treat the spectrogram as an image, and search for similar images using features at multiple scales, these approaches may be more prone to false detections or missed detections in the presence of strong narrowband noise. The dense binary Philips fingerprints are associated with Hamming similarity measure, so it requires a different similarity search implementation from that shared by the Waveprint and FAST fingerprints.

The fingerprint extraction method that demonstrates the most consistent performance at low SNR is the FAST fingerprints, a modified version of the Waveprint method, with wavelet coefficients standardized using the MAD statistics. The standardization step improves detection performance of the original Waveprint method because coefficient standardization induces a more uniform distribution of fingerprints in the continuous data set, with the effect of lowering the similarity between fingerprints corresponding to noise while maintaining high similarity between earthquake fingerprints. Although the standardization was originally implemented using the mean and standard deviation (Yoon et al. 2015), our recommendation is to use the median and MAD statistics for coefficient standardization in FAST fingerprints.

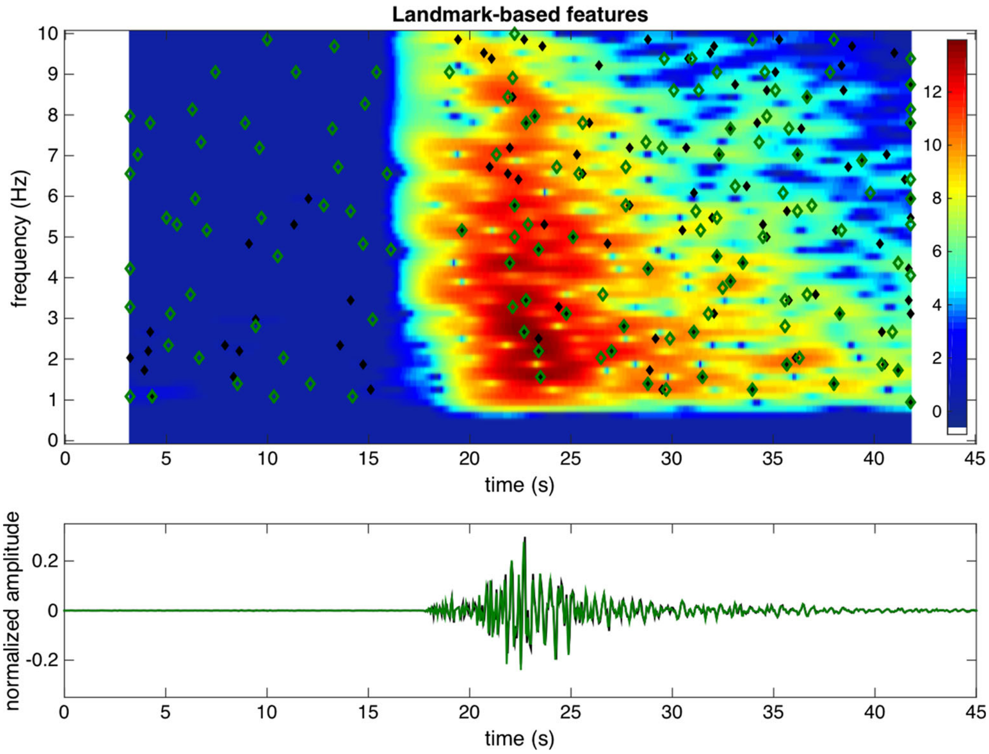


Figure 11

Illustration of the appropriateness of local maxima in the spectrogram as features for waveform similarity search for the earthquake detection problem. The lower panel shows two events with nearly identical waveforms, plotted in green (foreground) and black (background), and normalized to equal amplitude. The upper panel shows the spectrogram corresponding to the green event waveform, with the local maxima, or key points, identified by the landmark-based algorithm marked with green and black diamonds (corresponding to the green and black event waveforms, respectively). Although the green and black waveforms are nearly identical in the time domain, the key points for these two events have limited correspondence with each other. Local maxima in the spectrogram work well for music audio signals, which are typically localized in frequency and time, but these features are not as useful for identifying similar earthquake waveforms. Colormap is plotted on a log scale

6.1.1 Computational Concerns

The metric used to compare different feature extraction methods (Sect. 4.3) is a measure of how discriminative the resulting fingerprints are for waveform similarity search, but this measure does not capture other advantages or disadvantages of these methods with respect to our task. Our ultimate aim is to select a feature representation that will enable earthquake detection via blind search for similar earthquake waveforms in long-duration data. An important factor in selecting a feature extraction method is whether the chosen approach can be scaled to large data sets; the computational cost of feature extraction, the compactness of the feature representation, the effect on similarity search memory usage, and runtime must also be considered in selecting a fingerprinting scheme for use in waveform similarity search. The runtime for fingerprint extraction should not be prohibitive for data sets with durations on the order of months to years, though a slower feature extraction method that can be computed in a parallel or distributed manner would also be acceptable.

The need to estimate the median and MAD statistics for the distribution of each wavelet coefficient is a disadvantage of the FAST fingerprint extraction scheme from a computational standpoint. For large data sets, this can be partially addressed by computing the statistics on a sample of fingerprints rather than the entire data set. However, since the computational bottleneck in the FAST detection pipeline is often the similarity search step, it may be desirable to select a slightly slower feature extraction method if this choice of fingerprinting scheme will result in a speed-up in the similarity search runtime. The slight increase in runtime required to compute FAST fingerprints compared to Waveprint (due to the extra standardization step) can actually reduce the overall runtime if it results in fingerprints that are more uniformly distributed and produces fewer of the spurious matches that can significantly slow similarity search runtime.

Ten years of continuous data with a 1-s fingerprint lag corresponds to over 300 million fingerprints per channel. These fingerprints must be stored in memory to create the set of hash tables for the search index used in efficient similarity search. The fingerprint representation should be compact, requiring a relatively small number of bits to store each fingerprint. For FAST fingerprints of dimension $D = 4096$ with sparsity parameter $K = 400$, the storage requirement is $2 \times 4096 = 8192$ bits (binary representation) or $16 \times 400 = 6400$ bits (integer representation) per second of continuous data; this corresponds to 250–325 GB (gigabytes) to store fingerprints for 10 years of data. The Philips fingerprints have a computational advantage over the Waveprint and FAST fingerprints in this respect. While each Philips fingerprint requires $2 \times 32 \times 100 = 6400$ bits to store, there is significant overlap between the representation of adjacent fingerprints, so that in practice the storage requirement is only $2 \times 32 \times 10 = 640$ bits per second of continuous data, or 25GB for 10 years of continuous data, a factor of 10 reduction compared to the FAST fingerprints.

6.2. *Feature Learning for Waveform Similarity Search*

All of the fingerprinting methods discussed in this work are examples of "hand-engineered" features; they apply a set of fixed transformations that were selected based on expert knowledge of the properties of audio data and that have been demonstrated to perform well in practice. The three audio fingerprinting methods are data-independent, while the FAST fingerprints are modified to include an additional data-adaptive component. As an extension of the work presented in the previous sections, we investigated the use of fully data-driven approaches to feature extraction for waveform-similarity-based earthquake detection. Approaches that attempt to learn an optimal feature representation for a data set from the data itself are called *feature learning* or *representation learning* methods (Bengio et al. 2013) in machine learning. Learned feature representations have the potential to produce more discriminative fingerprints because they can be optimized for a given data set.

Feature learning encompasses a range of techniques, including matrix factorization (Deerwester et al. 1990; Lee and Seung 1999), dictionary learning (Lee et al. 2007), and neural networks (Hinton and Salakhutdinov 2006). We also considered a related set of methods called *learning to hash* (Wang et al. 2014, 2018), a data-dependent alternative to locality-sensitive hashing for approximate similarity search. For example, spectral hashing (Weiss et al. 2009) learns a set of data-dependent projections instead of using random projections as in locality-sensitive hashing.

There are a number of challenges that arise when applying feature learning or learning-to-hash methods to uninformed waveform-similarity-based earthquake detection (Bergen 2018).

It may be possible to learn high-quality features for similarity search using a labeled training set (collection of known earthquake waveforms). However, FAST is intended to be an uninformed method for earthquake detection. One of the properties that makes FAST particularly useful is that it can be applied in cases when there is a limited record of past seismic activity (i.e. the available catalog of template waveforms is limited or incomplete). Thus the feature learning approach should not require or assume the availability of known template waveforms; in the context of learning algorithms, this means limiting the approach to unsupervised methods, i.e. those that do not require labeled training data.

There are many unsupervised methods for feature learning and learning to hash that do not require template waveforms, but seismic data sets pose an additional challenge. Seismic data sets are imbalanced, meaning that the events of interest represent a minority of the signals in continuous data sets dominated by long periods of noise. Machine learning tasks with imbalanced data often result in poor performance on the minority class (in this case earthquake signals) because the learning algorithm has less information about the minority class, and the training samples from the minority class carry less weight overall in the learned model (He and Garcia 2009). As a result, it is difficult to learn a feature representation that produces low similarity between noise waveforms and high similarity between earthquake waveforms with unsupervised feature learning.

The challenges posed by the lack of labeled training data and the imbalanced, noise-dominated data make it difficult to provide FAST users with a single approach or framework for learning features that can be applied to diverse data sets. Therefore, our proposed FAST fingerprints, the data-adaptive variant of the Waveprint with MAD-standardization of the wavelet coefficients, are designed to be applied to any data set regardless of the availability of template waveforms, and take into account the data set imbalance problem. The coefficient standardization step in the FAST fingerprint extraction method is similar to the approach taken by some learning-to-hash methods: creating balanced hash codes that distribute the data more evenly in the hash tables (Weiss et al. 2009; Wang et al. 2010).

6.3. *Benchmarking for Performance Evaluation*

The approach outlined in Sect. 4 of this work can be used to evaluate alternative feature extraction schemes for waveform data beyond the audio fingerprinting methods presented in this work. Our framework for evaluating performance includes two key elements: benchmark data sets and an appropriate performance measure. The use of benchmark waveform data sets, containing known examples of noise and earthquake waveforms, provides a means of determining the detection performance for weak events. This allows us to overcome one challenge in evaluating the performance of algorithms for detecting weak events often missed in existing catalogs: the lack of ground truth, an objective standard for measuring detection performance. This framework provides a way to quantify the effectiveness of each feature extraction method for separating weak earthquake signals from noise in the context of similarity search. The experiments presented in this work only test for the distortions in the waveforms due to additive noise, but a similar framework can be used to compare fingerprinting schemes with respect to other criteria.

While our approach allows us to validate the feature extraction algorithm used in FAST, there remains the ongoing challenge of assessing detection performance in continuous data without ground truth. The goal of developing new earthquake detection algorithms is to outperform existing methods, but the nature of our task, discovering new events and sources, makes it inherently difficult to measure performance. Therefore, an open challenge in earthquake detection research is how to validate new candidate events, specifically how to distinguish true earthquake signals from transient signals due to local noise sources such as vehicles, transportation systems, and air traffic (Díaz et al. 2017; Meng and Ben-Zion 2018). The development of appropriate benchmark data sets and performance measures for comparing and validating earthquake detection algorithms will be critical for leveraging new developments in machine learning and data mining for seismology research.

Acknowledgements

This research is supported by National Science Foundation Grants EAR-1551462 and EAR-1818579. Computing resources were provided by the Center for Computational Earth and Environmental Science, Stanford University. Waveform data, metadata, or data products for this study were accessed through the Northern California Earthquake Data Center (NCEDC 2014). The ObsPy Python toolbox (Beyreuther et al. 2010) was used to retrieve and process waveform data.

Appendix

Signal-to-Noise Ratio

In this work, the signal-to-noise ratio (SNR) is computed using a 15-s interval of waveform data following the P-wave arrival. For signal x and noise

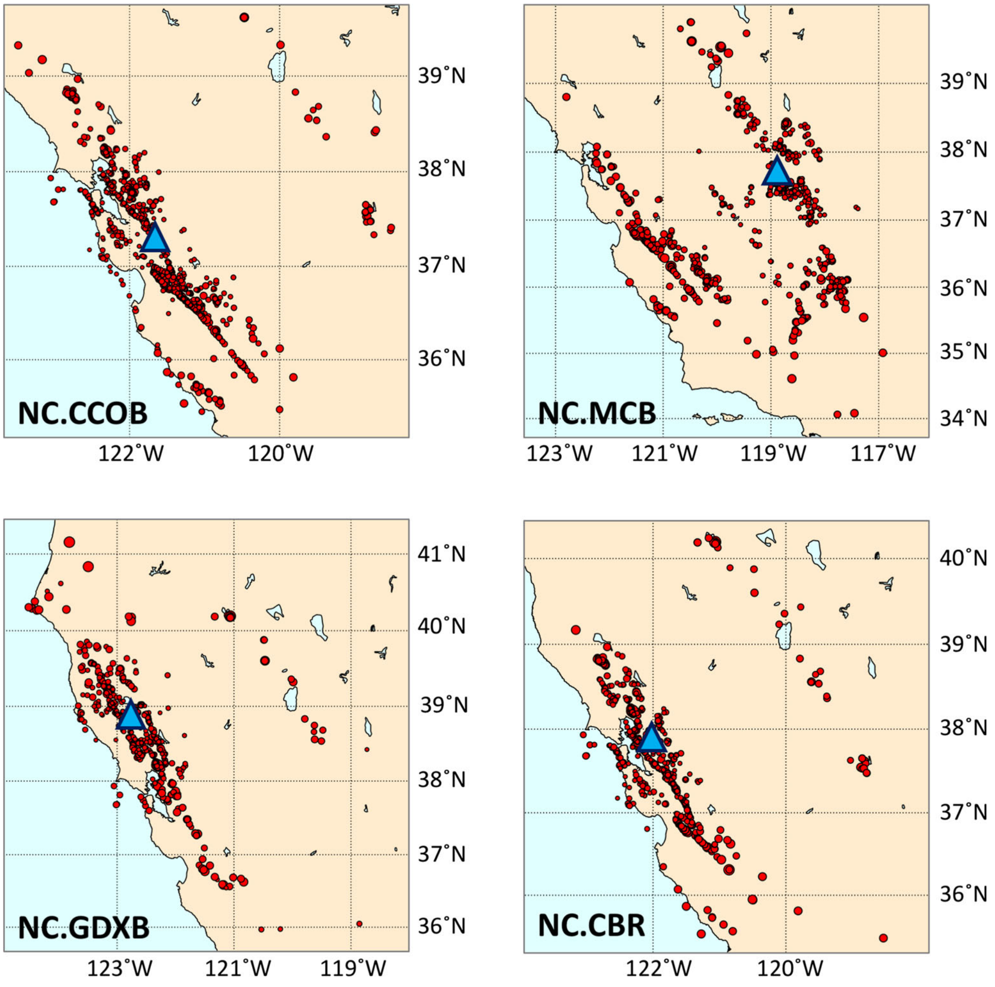


Figure 12
Map plots show the epicenter locations for the events in the earthquake waveform benchmark data set for stations NC.CCOB, NC.MCB, NC.GDXB, and NC.CBR. The location of each station is marked with a blue triangle, and the event epicenters are shown in red. The set of earthquakes is selected independently for each station depending on whether a P-phase arrival was recorded in the NCSN phase-pick catalog for a given event

n, each of duration $M = 300$ samples (15 s of data sampled at 20 samples per second), the SNR is given by:

$$SNR = \frac{P_{\text{signal}}}{P_{\text{noise}}}, \quad \text{where} \quad P_{\text{signal}} = \frac{1}{M}\sum_{i=0}^{M} |x[i]|^2, \text{ and } P_{\text{noise}} = \frac{1}{M}\sum_{i=0}^{M} |n[i]|^2. \tag{11}$$

Noise Segment Classification

Noise segments are assigned one of these three labels based on two criteria: (1) an STA/LTA threshold, and (2) a uniform energy criterion that quantifies how uniform the energy is across the 2-min interval. A noise segment is labeled as “clean” noise if the maximum STA/LTA ratio in the interval is below 3.0 and the uniform energy score is below 0.1. A noise segment is labeled as “non-noise” if the maximum STA/LTA ratio in the interval exceeds 6.0 or the uniform energy score exceeds 0.2. All other noise segments are labeled “lively” noise. The parameters for short and long windows used in the STA/LTA ratio are 3 and 45 s, respectively. The uniform energy score, $u^{(j)}$, associated with a noise segment $n^{(j)}$ of length M samples is defined as:

$$u^{(j)} = \frac{1}{M}\sum_{k=1}^{M}\left| s^{(j)}[k] - \frac{k}{M} \right|, \qquad \text{where} \quad s^{(j)}[k] = \sum_{i=1}^{k}\left(n^{(j)}[i]\right)^{2}, \tag{12}$$

and $n^{(j)}$ is normalized such that $||n^{(j)}||_2^2 = 1$. This value represents the difference between the cumulative signal energy over the interval and the cumulative energy for a signal with uniform energy, and u takes values between 0 and 0.5, with larger values associated with larger deviations from uniform signal energy.

Events in Benchmark Data Set, by Station

The list of events in the benchmark data set and the corresponding earthquake waveform and noise data matrices, by station, are available upon request (see Figs. 12, 13).

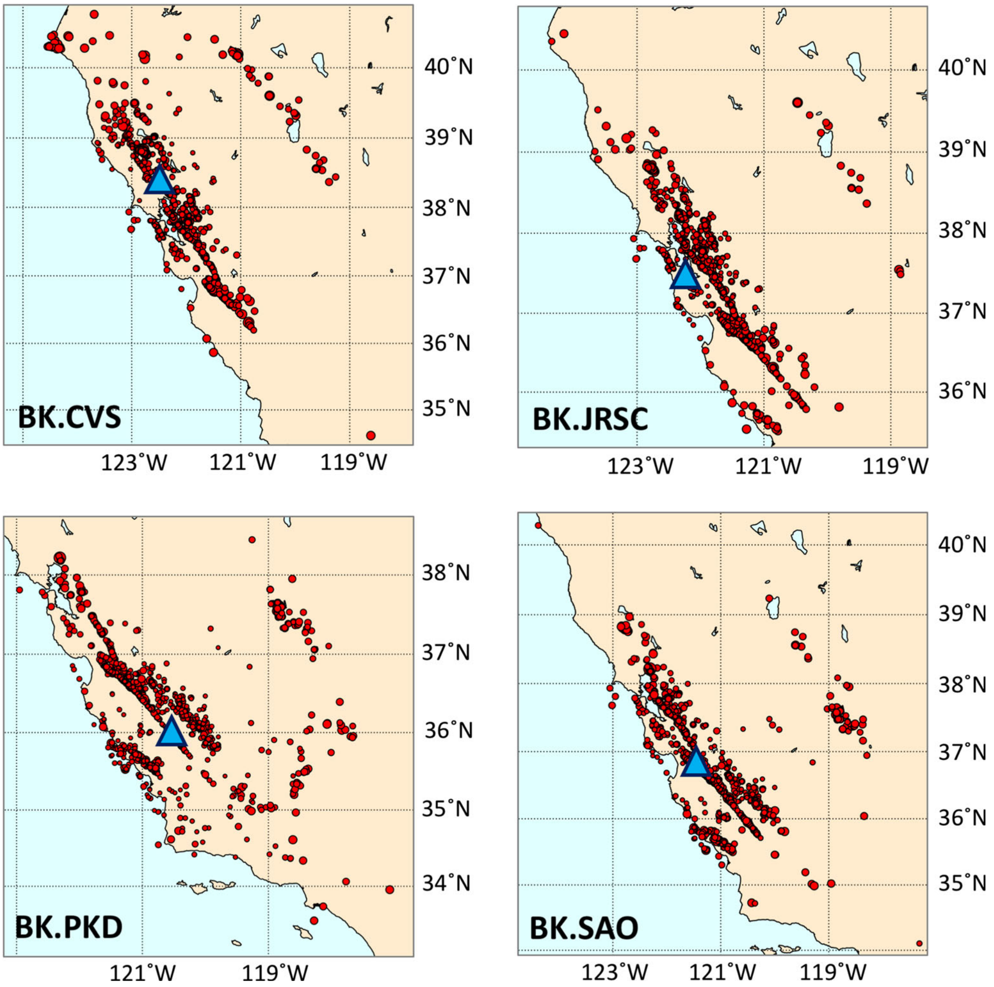


Figure 13

Map plots show the epicenter locations for the events in the earthquake waveform benchmark data set for stations BK.CVS, BK.JRSC, BK.PKD, and BK.SAO. The location of each station is marked with a blue triangle, and the event epicenters are shown in red. The set of earthquakes is selected independently for each station depending on whether a P-phase arrival was recorded in the NCSN phase-pick catalog for a given event

References

Alías, F., Socoró, J. C., & Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences*, *6*(5), 143.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, *72*(6B), S225–S242.

Andoni, A., & Indyk, P. (2006). Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, (pp. 459–468). IEEE.

Baluja, S., & Covell, M. (2008). Waveprint: Efficient wavelet-based audio fingerprinting. *Pattern Recognition*, *41*(11), 3467–3480.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509–517.

Bergen, K., Yoon, C., & Beroza, G. C. (2016). Scalable similarity search in seismology: a new approach to large-scale earthquake detection. In *International Conference on Similarity Search and Applications* (pp. 301–308). Springer, Cham.

Bergen, K. J. (2018). *Big Data for Small Earthquakes: Detecting Earthquakes over a Seismic Network with Waveform Similarity Search*. PhD thesis, Stanford University, Stanford, CA.

Bergen, K. J., & Beroza, G. C. (2018). Detecting earthquakes over a seismic network using single-station similarity measures. *Geophysical Journal International*, *213*(3), 1984–1998. https://doi.org/10.1093/gji/ggy100.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530–533.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159.

Broder, A. Z. (1993). Some applications of Rabin's fingerprinting method. In *Sequences II*, (pp. 143–152). Springer.

Broder, A.Z. (1997). On the resemblance and containment of documents. In *Compression and Complexity of Sequences 1997. Proceedings*, (pp. 21–29). IEEE.

Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2005). A review of audio fingerprinting. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, *41*(3 SPEC. ISS.), 271–284.

Datar, M., Immorlica, N., Indyk, P., & Mirrokni, V. S. (2004). Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the twentieth annual symposium on computational geometry*, (pp. 253–262). ACM.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391.

Díaz, J., Ruiz, M., Sánchez-Pastor, P. S., & Romero, P. (2017). Urban seismology: On the origin of earth vibrations within a city. *Scientific Reports*, *7*(1), 15296.

Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, *81*(3), 425–455.

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149–166.

Haitsma, J., & Kalker, T. (2002). A highly robust audio fingerprinting system. *Proceedings of the 3rd international society for music information retrieval conference (ISMIR02)*, (pp. 107–115).

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*(346), 383–393.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Science Advances*, *4*(5), eaao2929.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788.

Lee, H., Battle, A., Raina, R., & Ng, A. Y. (2007). Efficient sparse coding algorithms. In *Advances in neural information processing systems*, (pp. 801–808).

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of massive datasets* (2nd ed.). New York, NY, USA: Cambridge University Press.

Mallat, S. (2008). *A wavelet tour of signal processing: the sparse way*. Cambridge: Academic press.

Manber, U. (1994). Finding similar files in a large file system. In *USENIX Winter 1994 Technical Conference*, (pp. 1–10).

Meng, H. & Ben-Zion, Y. (2018). Characteristics of airplanes and helicopters recorded by a dense seismic array near Anza California. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1029/2017JB015240.

NCEDC. (2014). *Northern California Earthquake Data Center*. UC Berkeley Seismological Laboratory. Dataset. https://doi.org/10.7932/NCEDC.

Pankanti, S., Prabhakar, S., & Jain, A. K. (2002). On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(8), 1010–1025.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578.

Rong, K., Yoon, C. E., Bergen, K. J., Elezabi, H., Bailis, P., Levis, P., et al. (2018). Locality-sensitive hashing for earthquake detection: A case study of scaling data-driven science. *Proceedings of the VLDB Endowment*, *11*(11), 1674–1687.

Valentine, A. P., & Trampert, J. (2012). Data space reduction, quality assessment and searching of seismograms: Autoencoder networks for waveform data. *Geophysical Journal International*, *189*(2), 1183–1202.

Wang, A. (2003). An industrial strength audio search algorithm. In *ISMIR*, Vol. 2003, (pp. 7–13). Washington, DC.

Wang, A. (2006). The Shazam music recognition service. *Communications of the ACM*, *49*(8), 44–48.

Wang, J., Kumar, S., & Chang, S. -F. (2010). Semi-supervised hashing for scalable image retrieval. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, (pp. 3424–3431). IEEE.

Wang, J., Shen, H.T., Song, J., & Ji, J. (2014). Hashing for similarity search: A survey. arXiv preprint arXiv:1408.2927.

Wang, J., Zhang, T., Song, J., Sebe, N., & Shen, H. T. (2018). A survey on learning to hash. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1.

Weiss, Y., Torralba, A., & Fergus, R. (2009). Spectral hashing. In *Advances in neural information processing systems*, (pp. 1753–1760).

Willett, P., Barnard, J. M., & Downs, G. M. (1998). Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, *38*(6), 983–996.

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, *1*(11), e1501057.

(Received April 13, 2018, revised August 1, 2018, accepted September 10, 2018, Published online October 8, 2018)