Final Technical Report

USGS Award Number: G14AP00016
Title: Digitization of Analogue Seismograms for Improved Seismicity Constraints in the Northeastern United States

Miaki Ishii
Department of Earth & Planetary Sciences
Harvard University
20 Oxford Street
Cambridge, MA 02138
(617)384-8066
(617)384-8249 (fax)
ishii@eps.harvard.edu

Term: January 1, 2014 to December 31, 2014

Abstract

Two goals of this project were salvaging of analog seismograms recorded at the Harvard-Adam Dziewoński Observatory (HRV) between 1933 and 1953 and development of software that can be used to digitize these seismograms. Nearly 80% of the collection from the HRV station has been cleaned and scanned, and the images have been made available online at the Harvard Seismology web pages. The first version of a digitization software, DigitSeis, has been completed, and it differs from existing software in its design philosophy, allowing it to semi-automatically trace and combine vertically-offset time marks with the main segment of the seismogram. Another characteristic of this software is its attention to detail. The built-in algorithms correct for image distortion that is unavoidable due to age and analog nature of the seismograms, and assigns timing information that is as accurate as possible using the time marks. It has been successfully applied to a recording that contains multiple earthquake signals.

As a result of this project, an article describing the HRV collection and the procedures taken to rescue the recordings has been published in *Seismological Research Letters*. A manuscript has also been submitted to *Bulletins of Seismological Society of America* to report the development of DigitSeis and some of the algorithms that are implemented in this software. A step-by-step User's Manual for this software is also compiled. These three documents are submitted as parts of the Final Technical Report. Once the DigitSeis manuscript is accepted, it will be made available online along with its User's Manual.

Report

The main aim of the project was the development of software that requires as little human interaction as possible while delivering an accurate rendition of the analog seismograms in digital form. The other aspect of the project was continuation of efforts to clean and scann the collection of photographic recordings made at the Harvard-Adam Dziewoński Seismological Observatory (HRV) between 1933 and 1953. For both components of the project, the goals described in the original proposal have been met and they are described separately in the following sections.

Cleaning and Scanning of HRV Seismograms

This labour-intensive effort has been undertaken by our dedicated volunteer, Hiromi Ishii, and significant progress has been made during the term of the project. At the time of proposal submission, 18 boxes containing variable numbers of dirty seismograms had not been processed. As of March 1, 2015, this number is reduced to 6. We have chosen to proceed from the cleanest of the boxes to the dirtiest, so the seismograms in the last set of boxes require substantial amount of time to clean. The volunteer has been diligently and carefully restoring these seismograms, and has successfully recovered more seismograms than initially expected. We now estimate the amount of unrecoverable seismograms to be about 5% of the collection rather than 15% from the initial estimate. The volunteer has also continued to scan cleaned seismograms. At the time of proposal submission, almost 1400 seismograms from a time period between 1944 and 1946 had been scanned. As of March 1, 2015, nearly 9300 seismograms from a period between 1933 and 1952 have been scanned and made available online. This number corresponds to about 78% of the total photographic seismogram collection from the station, and only about 10% remains to be cleaned and scanned.

The status of the cleaning and scanning of these seismograms have been intermittently updated online as part of the Harvard Seismology Group web pages:

http://www.seismology.harvard.edu/HRV/status.html

The user can check which boxes have been cleaned, scanned or left untouched, and there are links to scanned images. All scanned seismograms up to March 1, 2015 have been made available through these pages, and anyone interested can download the .tif files (typical file size for the trace-side of the seismogram is about 1.6 GB while the back side with hand-written information is about 90 MB). We have also described the Harvard analog seismogram collection and reported the availability of the scanned images in a *Seismological Research Letter* publication (Ishii et al., 2015; PDF file submitted with the final technical report).

While these seismograms from the Harvard-Adam Dziewoński Seismological Observatory were examined and cleaned, we discovered a small collection predating the main collection. Prior to the building of the vault at Harvard, MA, where the 1933–1953 seismograms were recorded (the facility is still in use today), seismographs were deployed in the basement of the Harvard University Museum building (26 Oxford Street, Cambridge, MA). The operation from this earlier time had been known, but whereabouts of the recordings were unknown. The older seismograms discovered in the collection appear to be those made in Cambridge, MA, and are dated as early as 1917 (although we have not inspected these recordings fully, so there may be earlier recordings). These seismograms are distinct from those from Harvard, MA, in that the size of the photographic paper is small and the earliest of these are made by a needle scratching on a sooted paper. We have not cleaned or scanned these recordings from Cambridge, MA.

Development of Digitization Software

The funding for this project is mostly used to develop a software that is suitable for converting scanned images of seismograms into time series that can be easily used for modern analysis. In the proposal, we identified difficulties associated with the time marks as our main problem and that the new software will address the issue. The new software, DigitSeis, takes care of the time marks effectively, but while investigating features of the images and writing the software, we discovered additional complications that are typically not considered in other software, i.e., image distortion. The current version of DigitSeis corrects for image distortion almost automatically, and provides timing that is as accurate as possible. The software is described in detail (with figures) in a manuscript that has been submitted to *Bulletins of Seismological Society of America* (a preprint is submitted with this final technical report), but we will describe it briefly below.

DigitSeis is a MATLAB software that is written by Petros Bogiatzis. It takes advantage of image-processing or other routines that are either built into the MATLAB package or are publicly shared by various authors. The software is designed so that it will not run automatically but with the presence of a human analyst. This was a strategy we adopted intentionally, and it differs from other digitization software. We feel that aiming for a fully automated software for all seismograms is impractical considering that seismograms can be highly complex. Sometimes it is difficult even for a human being to trace the signal. Our goal for DigitSeis is to have automation wherever possible but allow for user interaction wherever needed. This approach allowed for simplification of some steps compared to existing software. For example, time marks have been difficult to handle in some software that tried to automatically trace them. In the end, the user is typically forced to trace individual time marks manually. By requiring user presence, DigitSeis can treat time marks semi-automatically, significantly reducing the amount of work required by the user. DigitSeis is also distinguished by its painstaking efforts to obtain as accurate time series as possible. This includes analyses for often-ignored features such as distortion of the images, and some of these features are highlighted below.

The main motivation for the development of DigitSeis was the challenge of working with seismograms where vertical offsets are introduced for a short time to indicate minute or hour point (time marks). In order to effectively treat the time marks, the DigitSeis software has a classification step where different objects within the image are grouped into time marks and main (albeit disconnected) traces. This is performed semi-automatically using the vertical separation between the time marks and main traces with some user input for poorly separated or strange objects (e.g., stains and handwritten notes). Once the classification is finished, the time mark objects are amalgamated with the main trace with varying level of vertical offset correction. The resulting continuous trace is checked using its derivative (if the vertical correction is insufficient, the derivative will be large), and the optimal vertical correction is determined. Because the software works under the assumption that these vertically-offset segments exist, the time marks are no longer an issue in the digitization procedure.

One property of the seismogram images that came as a surprise is the level of distortion they contained. On hindsight, this is not surprising considering the amount of curvature and deformation that exist in the original analog seismograms. The process of acquiring a digital image of these seismograms introduces additional distortion such as due to lens effect for images taken with digital camera (even for a wide-angle lens). Regardless of the source of distortion, it is often enough to jeopardize the tracing exercise by obscuring the zero-amplitude position and complicating the conversion of horizontal position to time. DigitSeis recognizes these issues and corrects for them, first by using relatively a simple model. This approach works well for determining information about the "vertical" position of the seismogram, but the model is too simple in the "horizontal" direction. Small yet significant variations in the length to timing occurs along the "horizontal" direction, and acquisition of accurate timing requires more sophisticated analysis. DigitSeis utilizes the positions of the time marks, and an algorithm determines the time associated with each time mark and interpolates between them to obtain means to assign time to a given "horizontal" location. This procedure, in addition to the distortion corrections, results in uneven time sampling. The software takes the largest time factor and generates an evenly sampled time series in SAC or MATLAB file format.

The biggest challenge for any software that attempts to digitize analog seismograms is crossing of traces when large disturbances (such as earthquakes) are recorded. There probably is no good way to automatically treat crossing seismograms since some decisions become subjective. DigitSeis does not attempt to unravel the traces, but relies on the user to perform this procedure. This is possibly the most time-consuming operation during a digitization session, and there is a room for future improvements.

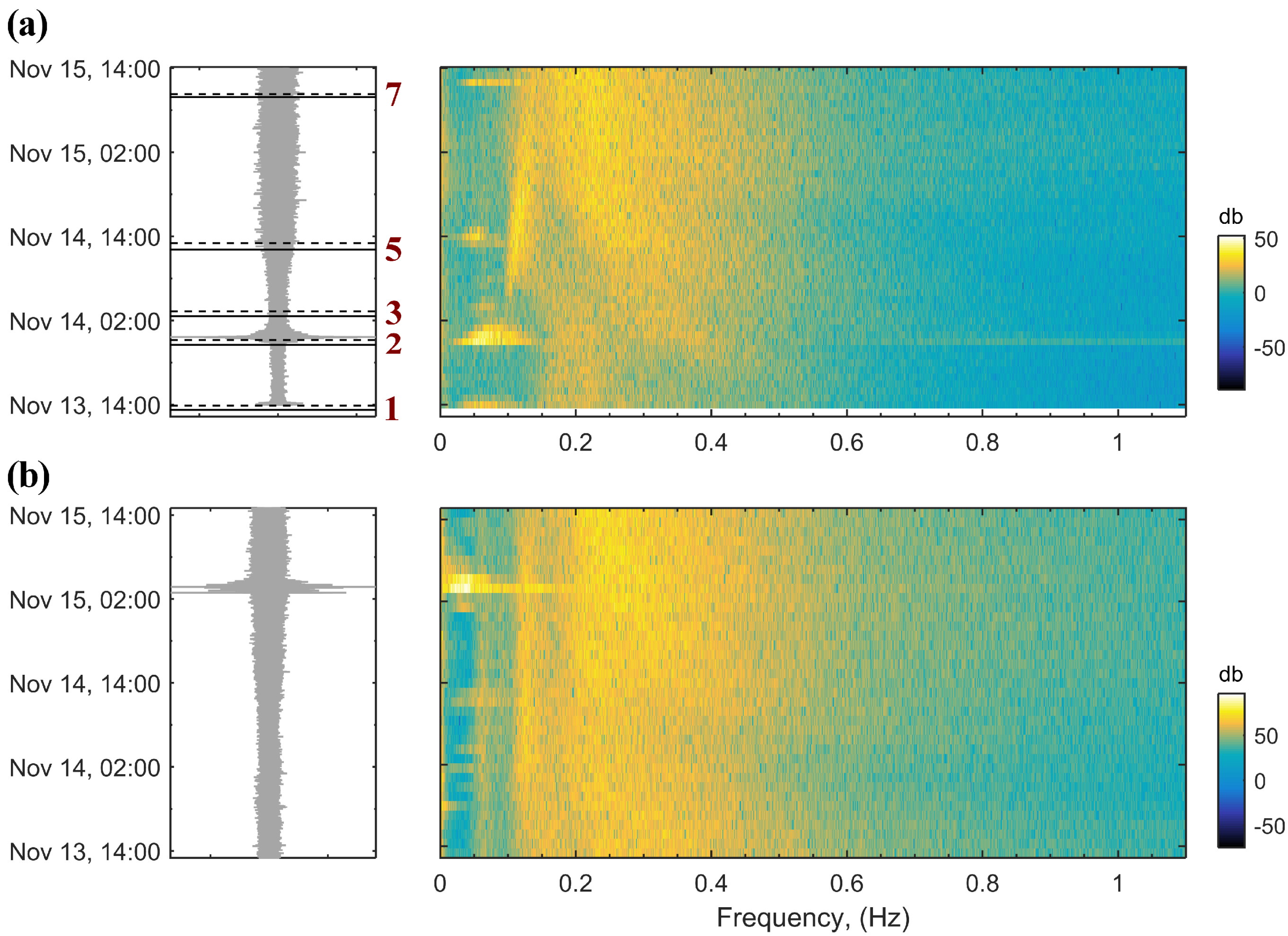

Figure 1: Comparison of digitized seismogram from 1938 with modern digital recording from 2014. (top left) Time series between November 13 and 15, 1938 obtained from application of DigitSeis to a long-period north-south recording. The solid horizontal lines show hypocentral times of some earthquakes that are reported in the International Seismological Center catalogue, and the dashed lines are the expected arrival times of the surface waves. The numbers beside these lines indicate different earthquakes. “1” is for an M_S 6.9 earthquake in the Kuril Islands region, “2” is an M_S 7.0 event in Japan that is accompanied by another event (“3”), “5” is an event in New Britain, Papua New Guinea, and “7” is an earthquake in the Aleutian arc.
(top right) Spectrogram of digitized seismograms.
(bottom left) Time series from similar days of the year as above except from year 2014. This time series also contains earthquake signals, and an M_w 7.1 event that occurred on November 15, 2014 is clearly visible.
(bottom right) Spectrogram of digital recording from 2014.

The DigitSeis software has been extensively tested, but since the focus of the project was on development rather than its application, it has not yet been systematically used to digitize the HRV seismogram images. We have, however, digitized a couple of seismograms to investigate the effectiveness of the software, and one example, a long-period north-south recording from 1938 is shown in Figure 1. The time period covered by the seismogram (November 13–15, 1938) had a number of large earthquakes, and the largest is clearly visible in the analog image (http://www.seismology.harvard.edu/HRV/HRV1938/s1677_1938_1113_1115_lp_ns_f.tif). When this seismogram is digitized, signal associated with another earthquake that occurs at the beginning of the time window becomes evident (Figure 1). With the seismogram in digital format, it is straightforward to apply some basic analysis, and its spectrogram is calculated. In this representation of the seismogram, additional earthquakes emerge. Signals of these events are not easily discernible on the raw seismogram, and it demonstrates the power of having the analog image in the digital format. Another notable feature of the spectrogram is the absence of signal at any particular frequency. If the algorithm that combined the minute time marks with the main trace is defective, we expect a peak at a frequency corresponding to a minute period (i.e., around 0.017 Hz). The absence of such feature confirms that DigitSeis is successful in digitizing seismograms without introducing spurious signals. Finally, there is unquestionable increase in the noise level around November 14, 1938. This noise has distinct frequency content with peaks at about 0.14 and 0.25 Hz. These features are consistent with those of noise recorded by modern instruments in 2014 (Figure 1) including dispersion as the source of noise arrives at the station. This comparison further illustrates the success of digitization and the potential of using the old analog seismograms for analyses of historical storm activities and how they may have changed over time.

In addition to the manuscript describing the features of DigitSeis that has been submitted to *Bulletins of Seismological Society of America*, Petros Bogiatzis has prepared User's Manual that describes step-by-step procedure for using the software (submitted with this final technical report). It is our intension to make the software and the Users's Manual available online at the Harvard Seismology web page (http://www.seismology.harvard.edu) once the BSSA paper is published.

Bibliography

1. Bogiatzis, P., & Ishii, M., 2015. DigitSeis v0.5: User's Manual. submitted with this technical report, and will be made available at Harvard Seismology Web page once the paper below is published.

2. Bogiatzis, P., & Ishii, M., 2015 (submitted). DigitSeis: A new digitization software for converting analog seismograms. *Bull. Seismol. Soc. Am.*

3. Ishii, M., Ishii, H., Bernier, B., & Bulat, E., 2015. Efforts to recover and digitize analog seismograms from Harvard-Adam Dziewoński Observatory. *Seism. Res. Lett.* **86**(1), 255-261.