Title
MyShake – Building a global smartphone seismic network

Author
Kong, Qingkai

Publication Date
2018

Peer reviewed|Thesis/dissertation

MyShake – Building a global smartphone seismic network

By

Qingkai Kong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Earth and Planetary Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Richard Allen, Chair
Professor Douglas Dreger
Professor Jonathan Bray

Spring 2018

MyShake – Building a global smartphone seismic network

Abstract

MyShake – Building a global smartphone seismic network

by

Qingkai Kong

Doctor of Philosophy in Earth and Planetary Science

University of California, Berkeley

Professor Richard Allen, Chair

Earthquakes are global hazards that account for many deaths and economic losses each year. With the development of seismology and technology, we have denser and denser sensors to monitor the earthquakes and provide valuable dataset to understand the processes of the plate tectonics, earthquake rupture processes, etc. Many useful applications built on top of these sensors including earthquake early warning systems that have the great potential to reduce the earthquake hazards for human civilizations. From the seismological point of view, the seismologists have been trying different ways to increase the density of the sensor network to monitor the earthquakes. From high-quality broadband seismic stations, to low-cost microcontroller devices, the instruments used in seismology are crossing the whole spectrum. In this dissertation, I report our progress on building a global seismic network using the consumer smartphones. The goal is to use the power of crowdsourcing devices to setup a scalable seismic network to compliment the existing high-quality seismic network, especially provide useful data at places where they cannot afford the high-quality instruments.

This thesis starts with the design of the methodology and experiments we did before the building of this global network, including the noise floor tests, shake table tests, and the design of the artificial neural network to distinguish the earthquake signals recorded on the phone from the daily human activities. With the earthquake early warning application in mind, these form the basis and the blue prints to build the network. In Feb 12th 2016, MyShake application and the whole system released to the public. Within very short time MyShake users cover 6 continent and starting to provide the shaking data related to the earthquakes. The initial observations from this network validate the initial design and concepts, at the same time it shows great potential to use the recorded data to do routine seismological applications. The two types of data from MyShake, i.e. real-time trigger messages and the 5-min long 3-component waveforms have different applications. The real-time trigger messages enable MyShake network to be used as a stand-alone earthquake early warning system, including estimate the initial location, magnitude and origin time of the earthquake. On the other hand, with the waveforms we recorded from the smartphones, we could refine these earthquake parameters at better accuracy. The comparison of the estimated locations, origin times and magnitudes from the MyShake recordings with those from the catalog shows the data

from these consumer devices are useful to quantify the earthquakes. This will be really useful by providing extra data at places where no or few seismic stations near the earthquake but with a large population. This thesis also talks about the potential application of using MyShake to conduct structural health monitoring of buildings in the future. The shaker test we did proves the sensors in the phones could be used to extract the fundamental frequencies from the shaking of the buildings.

The success of building this global smartphone seismic network and the initial analysis we conducted using the data recorded on it provide the community an exciting way to monitor earthquakes, though there are still many challenges and limitations need to be addressed. The last part of this thesis talks about the pathway forward from our experience with this network. Besides, MyShake project is an example of the combination of data science and earth science, I will end the thesis with a discussion of my thinking of how to take the full advantage of both sides.

Contents

Chapter 1 Introduction **1**

Chapter 2 MyShake methodology **4**
2.1 Abstract **4**
2.2 Introduction **4**
2.3 Results **6**
2.4 Materials and Methods **14**
2.4.1 Data collection 14
2.4.2 Noise floor test 14
2.4.3 Shake table test 14
2.4.4 Single phone detection algorithm design 14
2.4.5 Network detection algorithm design: 15
2.5 Discussion **15**
2.6 Supplementary material **16**
2.6.1 Data collection: The MyShake application 16
2.6.2 Classifier analysis: Detecting earthquakes on a phone 18
2.6.3 Network detection algorithm 23
2.6.4 Estimate warning time for Katmandu, Nepal 26
2.7 Acknowledgments **26**

Chapter 3 Initial Observations from MyShake Network **27**
3.1 Abstract **27**
3.2 Introduction **27**
3.3 MyShake Methodology Refresh **28**
3.4 Seismic data recorded by MyShake **30**
3.5 Discussion and conclusions **33**
3.6 Supplementary material **34**
3.7 Acknowledgments **48**

Chapter 4 Characterize the earthquakes **49**
4.1 Abstract **49**
4.2 Introduction **49**
4.3 Data used **51**
4.4 Timing and location accuracy of smartphone records **54**
4.5 Location and origin time estimation **55**
4.6 Magnitude estimation **61**
4.7 Discussion **64**
4.8 Conclusion **65**
4.9 Supplementary Figures **65**
4.10 Data and Resources **71**
4.11 Acknowledgements **71**

Chapter 5 The potential of using smartphones for structural health monitoring **73**
5.1 Abstract **73**
5.2 Introduction **73**

5.3 Background of Millikan Library **76**
5.4 Method **76**
5.5 Results **79**
5.6 Estimation of the orientation of the phones **82**
5.7 Discussion and Conclusions **84**
5.8 Supplementary material **86**
5.8.1 Structural health monitoring infrastructure 88
5.9 Acknowledgements **89**
Chapter 6 Conclusion **90**
Bibliography **94**

List of Figures

Figure 2.1 Noise floor of the phones. Noise floors of the smartphones color coded by the phone release date (also shown in the legend as MM/YY). Dashed black lines are typical ground motion amplitudes of earthquakes 10 km from the epicenter for various magnitudes. Noise floor for high quality MEMS sensor (HP MEMS - blue) and a typical force-balance accelerometer from a regional network (BKS in northern CA - purple) are also shown.....6

Figure 2.2: Three-dimensional shake table test. The input seismogram is from a real earthquake that has been modified for IEEE-693-2005 tests. (A) Waveform comparison between phone (blue) and reference accelerometer (red) recordings from an input signal that has peak acceleration of 0.5g. (B) Spectrum comparison of Y components. The X and Y components are in the plane of the phone, which is lying flat on the horizontal shake table and is not attached. The Z component is perpendicular to the plane of the phone and is vertical for this test.....7

Figure 2.3: Shake table test with an input sweep signal (0.5-7Hz). (A) Waveform comparison between a phone fixed on the table (blue), a phone placed freely on the table (black) and the reference accelerometer attached to the table (red). (B) Frequency domain comparison of the signals in (A). (C) Calculated correlation coefficient and RMS (Root Mean Square) ratio between the signal recorded by the phone placed freely on the shake table and the reference accelerometer. The correlation coefficient is a measure of the phase match and RMS is a measure for amplitudes match. We use 1 Hz frequency band to filter the record and calculate the coefficient with a step frequency 0.1 Hz. The x-axis is the center frequency of the frequency band. The correlation coefficient shows how well the phase is recorded by the phone, and the RMS ratio shows the amplitude recovery. Above 2-3 Hz the phone starts to slide so the full amplitude is not recovered, however, the phase is recovered up to 7-8 Hz.....8

Figure 2.4: Earthquake recorded by phone and classifying earthquakes. (A) Example 12-hour 3-component acceleration record from a private/personal Samsung Galaxy S4 phone starting at 4 pm August 23, 2014. It shows the accelerations of every-day human motions for the first ~8 hours, then appears stationary during the night. The red box at the end of the figure highlights the time window of figure b. (B) 1 minute of data from the period shown in (A) at the time of the M6 Napa earthquake 38 km from the phone. The earthquake occurred at 3:20:44 am local time. (C) Scaled feature plot showing IQR versus ZC for the classifier training dataset. The blue dots are the centroids of human activities, and the red dots are the earthquake features. (D) 3D plot of the 3 features we used to distinguish earthquakes. Adding the CAV to IQR and ZC drags some of the human activates (blue dots) to the third dimension but not the earthquake data, this helps improve the results. 9

Figure 2.5: Estimated magnitude. Comparison of our estimated magnitudes with the real magnitude for earthquakes in Japan using phone-like data. The green line is the 1:1 line, and the two grey lines are the 1 magnitude unit shift, each blue point is the magnitude estimate at a single simulated phone. The red pluses are the average event estimates, which is the average of multiple single phone estimates. 12

Figure 2.6: Snapshots of trigger detections for the 2014 M5.1 La Habra earthquake simulation at 3, 5 and 7 sec after the event origin time. Grey dots are stations; pink indicates a trigger. The true

earthquake location is the red star with circles at 10, 20 and 30 km radius. The blue star represents the estimated event location, first detected at 5 sec. The magnitude estimate at each point in time is shown upper right. 13

Figure 2.7: MyShake activity November 1, 2014 - February 28, 2015. (a) Number of phones that downloaded MyShake and registered with our network (green curve), and the number of active phones running the application on a given day based on the SOH information (blue curve). Server at CPC restarts during the first month is the reason the number of active phones drops to zero. (b) Number of phone triggers each day with waveforms uploaded to the CPC, a total of 17600 triggers were collected. 17

Figure 2.8: Example earthquake record used to train the ANN classifier algorithm. The waveform is the EW component from a regional network station 16.5 km from the epicenter of the western Tottori earthquake (M7.3) of October 6, 2000. The data has been modified to represent a smartphone recording at the same location. Only 2-sec windows of data from the yellow region were used to train our algorithm. 19

Figure 2.9: Structure of Artificial Neural Network (ANN) classifier algorithm. It has three layers: one input layer with 3 nodes, a hidden layer with 5 nodes, and an output layer with 1 node. For the hidden layer and output layer, the inputs from the previous layer to the each node will be first summed and then fed into an activation function shown as f. 21

Figure 2.10: Receiver operating characteristic (ROC) curve. Shows the ANN classifier performance on 30% test data split from the training data. The ROC curve shows the false positive rate (classified as earthquake when it is a non-earthquake) on the x-axis, , against true positive rate (classified as an earthquake when it is an earthquake) on the y-axis. Ideally, the curve will climb quickly toward the top-left corner meaning the model correctly predicted the cases. Our result is quite close to the ideal cases 22

Figure 2.11: Phone trigger times versus epicentral distance. The regional network data from California and Japan was modified to phone-quality data and then our classifier applied to the data to determine when a trigger occurs. The red line is the best-fit to the data and has a moveout velocity of 3.2 km/sec; most triggers are generated by the S-wave or the later surface wave. The blue outline is the time-space window used for association of triggers with an event by the network detection algorithm. 23

Figure 3.1: Distribution of MyShake registered users and detected earthquakes. (a) Registered MyShake users are shown in clusters. The number in each circle indicates the number of registered users in the cluster, and the color of the circle shows the order of the number of phones, i.e., a purple circle indicates the number of phones is of order tens of thousands, magenta is thousands, red is hundreds, yellow is tens, and blue for less than 10. (b) 237 Earthquakes recorded by MyShake users since February 12, 2016. The locations of the earthquakes are shown as circles, which are color-coded by the depth and whose sizes are scaled by the magnitude of the earthquake. Figures are generated on August 11th 2016. 29

Figure 3.2: (a) Location of the M5.2 Borrego Springs earthquake and the MyShake phones at the time of the event. Blue star is the epicenter of the earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were

ready to detect the earthquakes but did not. (b) MyShake trigger time vs distance. Blue dots are the phones' trigger times, and the green and red curves are the estimated P and S wave travel time based on Model ak135 [*Kennett et al.*, 1995]. 31

Figure 3.3: (a) Record section plot for phones within 200 km. Each blue trace is one horizontal recording from MyShake user, and the green and red curve is the estimated P and S wave based on ak135. Amplitudes of the recordings are normalized in each trace. (b) PGA value observations with distance. PGA values from MyShake (blue) and traditional seismic stations (red) are shows as observed on the largest horizontal component. The seismic station data are from Southern California Earthquake Data Center (SCEDC)..... 32

Figure 3.4: Example MyShake waveforms. (a and b) Comparison of the waveforms recorded by MyShake and a nearby traditional seismic station (horizontal component) for the M5.2 Borrego Springs earthquake. (a) MyShake waveform recorded 37.2 km from the epicenter, and a traditional seismic station 0.88 km from the smartphone. (b) MyShake waveform recorded at 100.9 km from the epicenter, and a traditional seismic station 1.93 km from the smartphone. See Figure 3.10 and Figure 3.11 for the comparison of other components. (c) M7.8 Ecuador earthquake recorded at 170 km away from epicenter. (d) M5.1 Oklahoma earthquake recorded by a phone at 130 km from epicenter. For (c and d) the zero time is the phone trigger time. The vertical black, green and red lines are the origin time and predicted P- and S-wave arrival time respectively (estimated using ak135 model)..... 32

Figure 3.5: The time history of the phones registered with our server (blue curve), and the phones actively contributing data to MyShake in a 24-hour interval (green curve). Registered phones are defined as the phones with MyShake installed that have sent at least one data point to our server. The total number of registered phones is a little lower that the total number of phones that have downloaded and installed the app. Active phones are defined as the phones sent data back to the server within last 24 hours..... 34

Figure 3.6: Waveform examples from different parts of the world recorded by MyShake. The earthquake location/time is shown. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 min buffer before the trigger. The green and red lines are the estimated P and S arrival times using the ak135 model. The waveform from Nepal has about 1 second missing data, which occasionally happens in MyShake app. Continues below..... 35

Figure 3.7: Magnitude and depth distribution of MyShake detected earthquakes. 38

Figure 3.8: Record section plot for the vertical component of MyShake recorded waveforms. The green and red lines are the estimated P- and S-arrival times estimated using ak135..... 39

Figure 3.9: PGA comparison and observation times. (a) Histogram of PGA difference between MyShake recordings and the closest traditional seismic station. A positive value means that the PGA value observed by MyShake is larger than that from the traditional seismic station. (b) PGA observation times for MyShake and traditional seismic stations. The blue and red curves are the estimated P- and S-wave arrival using ak135..... 40

Figure 3.10: Waveform comparisons of all 3 components for a smartphone 37.2 km from the M5.2 Borrego Springs earthquake with a traditional seismic sensor 0.88 km away. The blue traces

are recorded by MyShake phone, and the red traces are recorded by the seismic station. Both the raw waveforms and filtered waveforms are shown.....41

Figure 3.11: Waveform comparisons of all 3 components for a smartphone 100.9 km from the M5.2 Borrego Springs earthquake with a traditional seismic sensor 1.93 km away. The blue traces are recorded by MyShake phone, and the red traces are recorded by the seismic station. Both the raw waveforms and filtered waveforms are shown.....42

Figure 3.12: Map of the M7.8 Ecuador Earthquake and one 3-component recording. (a) Location of the 2016-04-16 M7.8 Ecuador Earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were ready to detect the earthquakes but did not. (b) 3 component recordings from a user at 170.3 km from the earthquake. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model. The large amplitudes after the S wave arrival are likely from the user picking up the phone when he/she felt the earthquake. The ground motion recorded before the human activities are shown in Figure 3.4c. 43

Figure 3.13: Map of the M5.1 Oklahoma earthquake and one 3-component recording. (a) Location of the 2016-02-13 M5.1 Oklahoma Earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were ready to detect the earthquakes but did not. (b) 3 component recordings from a user at 130.5 km from the earthquake. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model. 44

Figure 3.14: Additional examples of P-waves recordings for larger earthquakes. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model.....45

Figure 3.15: Top figure shows an example of the signal-noise-ratio (SNR) as a function of frequency for the horizontal component in Figure 3.10. It is computed by using the amplitude of the FFT of the signal divide by the mean noise spectrum amplitude. The phone is at 37.18 km away from the M5.2 earthquake in Southern California. The bottom figure shows time frequency representation of the signal, including frequency spectrum, spectrogram, and the time domain waveform for this recording.....46

Figure 3.16: Top figure shows an example of the signal-noise-ratio (SNR) as a function of frequency for the horizontal component for the 2016-05-18 07:57:05 UTC, M6.7 Ecuador earthquake. The phone is at 99.41 km away from the earthquake. The bottom figure shows time frequency representation of the signal, including frequency spectrum, spectrogram, and the time domain waveform for this recording.....47

Figure 4.1: (a) Earthquakes with one or more useful waveform recordings from MyShake phones in the first two years of operation (Feb 12, 2016 to Feb 12, 2018). The size of the circle and color

represent magnitude and depth of the earthquake (both magnitudes and locations are from USGS ComCat catalog). The small inset figure shows the density of the MyShake users with warmer color showing more users. (b) The heatmap of the number of unique MyShake phones within 3-degree bins. (c) Distribution of earthquake magnitudes for which one or more useful waveforms were recorded.by MyShake (blue triangles) and the catalog events (red circle). The number of events is measured in 0.5 magnitude bins. Figure 4.11 are showing the difference of the number of events recorded by MyShake and the USGS catalog on the map. 51

Figure 4.2: Distance of waveform recordings recorded by MyShake as a function of magnitude. The blue dots are waveforms recorded by MyShake, and the red curve is the fitted line to the furthest recordings from MyShake database. We only search M2.5 and above events that are corresponding to the USGS catalog event using a space and time window. The red curve is the analytic representation for estimate the furthest waveforms we expect to see for different magnitudes. The inset figure on the top left is the cumulative distribution of the signal to noise ratio (SNR) for all the earthquake recordings measured on one horizontal component (Y component). 52

Figure 4.3: 3-component acceleration waveforms from MyShake detections globally. The black line is the event origin time from USGS catalog, green and red lines are estimated P and S arrival time from model ak135 [*Kennett et al.*, 1995]. Time 0 on each panel is the time when the phone thinks it detects an earthquake. 54

Figure 4.4: The location of the 69 selected events with more than 5 useful waveforms for anlysis. Size of the circle represents the magnitude and color coded by the depth of the event. 56

Figure 4.5: (a) The distribution of the epicentral distance errors from the estimation using the time difference between P and S phases. The red line is showing the mean value. There are total of 18 events. (b) The distribution of the corresponding 18 events epicentral distance errors from the estimation using S phase pickings alone. The red line is showing the mean value. 57

Figure 4.6: (a) Estimated origin time error vs the epicentral distance error using only the S phase pickings. The size of the circle is showing the number of waveforms used in the calculation and the color is showing the station coverages that estimated from the azimuth of the two furthest stations and check whether it is in the range of 90 degree, 180 degree, 270 degree, and 360 degree as an indicator of the station coverage. The mean epicentral distance error is 17.8 km, median error 6.1 km and with a standard deviation 22.8 km. The mean origin time residual is 0.1 s (Estimate origin time – Catalog origin time), median error is 1.0 s, and standard deviation 6.9 s. (b) The histogram of the epicentral distance error from only the best coveraged events which have the phones' azimuthal range larger than 270 degree. The red line is the mean value 5.5 km, the median is 2.7 km and standard deviation is 7.2 km. 58

Figure 4.7: Location estimation example from a well-recorded event - M4.4 event in Berkeley on 2018-01-04, 10:39:37.730 (UTC). The magenta dots are the location of the waveforms used in the analysis. The red star is the USGS catalog location and the blue star is the estimation location. The error in the epicentral distance is 1.5 km and with a 2 s origin time error. The grid-search is starting from the whole region showing in the map and finds the best solution at a coarse grid. The grid dots showing here are the finer grids in the analysis that color-coded by the sum of the square of the difference between the observed S phase time and the estimated S phase. 59

Figure 4.8: (a) Location estimation example from M4.6 event in Peru on 2017-02-01, 11:38:30.100 (UTC). The magenta dots are the location of the waveforms used in the analysis. The red star is the USGS catalog location and the blue star is the estimation location from the finer grid search using only S phase pickings. The epicentral distance error is 64.8 km between the estimated and catalog location. The grid-search is starting from the whole region showing in the map and finds the best solution at a coarse grid. The grid dots showing here are the finer grid in the analysis that color-coded by the sum of the square of the difference between the observed S phase time and the estimated S phase (b) Location estimation for the same events but using both P and S phase pickings. The location error is 23.2 km. 60

Figure 4.9: (a) Estimated magnitudes versus USGS catalog magnitudes using the estimated locations. The red line is 1 to 1 magnitude, and the two dashed lines are 1 magnitude unit off from the 1 to 1 line. Each dot is estimated using different number of waveforms illustrated by the color. The inset plot is histogram of the residual magnitude (estimated magnitude – catalog magnitude). The mean of the residual is 0.3, and standard deviation is 0.4. The line in the histogram is the mean value. (b) Estimated magnitudes versus USGS catalog magnitudes using the estimated locations after waveform amplitude correction by dividing the amplitude by a factor of 1.6 that found by best matching the USGS catalog magnitude. The mean of the residual is 0.1, and standard deviation is 0.4. 62

Figure 4.10: (a) Estimated magnitudes versus USGS catalog magnitudes using the USGS catalog locations. The red line is 1 to 1 magnitude, and the two dashed lines are 1 magnitude unit off from the 1 to 1 line. Each dot is estimated using different number of waveforms illustrated by the color. The inset plot is histogram of the residual magnitude (estimated magnitude – catalog magnitude). The mean of the residual is 0.2, and standard deviation is 0.3. The line in the histogram is the mean value. (b) Estimated magnitudes versus USGS catalog magnitudes using the USGS catalog locations after waveform amplitude correction by dividing the amplitude by a factor of 1.6 that found by best matching the USGS catalog magnitude. The mean of the residual is 0.0, and standard deviation is 0.3. 63

Figure 4.11: (a) The ratio of number of MyShake recorded events over that from USGS catalog within 3-degree bins. It includes all events above M2.5. The colors are showing the ratio. (b) Same as (a), but it shows all the events above M5.0 using a 5-degree bin. 66

Figure 4.12: Location errors from the smartphone GPS experiments. The error bars are showing the standard deviations from all the measurements for each phone. 67

Figure 4.13: Timing error for the phones within 1 hour estimated by the offsets from the NPT synchronization. The red line is showing the median of the data and the box showing the 25 and 75 percentile, while the two bars at the top and bottom are whiskers which is 1.5 times of the Interquartile Range (IQR) from the box. 67

Figure 5.1: Background of the test. (A) The Millikan Library building viewed from the northeast. The two dark colored panels on the near-side of the building comprise the east shear wall (modified from [Bradford et al., 2004]). The inset figure (modified from [Clinton et al., 2006]) shows the plan of the building, where the star is the location of the test phones and the dot is the location of the Episensor station CI MIK. (B) The shaker located on the roof, used to generate oscillation of the building. The two exposed buckets contain lead masses which spin in opposite directions to generate a sinusoidal horizontal force. (C) The 25 smartphones used

in the test, all placed on the floor of the 9th (top level) floor. The duration of the shaking in the North-South direction is 1:29 pm to 3:02 pm, and in the East-West direction is 3:38 pm to 5:03 pm, local time. 75

Figure 5.2: Waveform comparisons. Waveform comparisons between the Episensor (CI MIK) and a Samsung Galaxy S4 phone for the two horizontal components. (A) North-south component. (B) East-west component. The red time series is from the single phone recording (Samsung Galaxy S4), and the blue is from the MIK recording. Frequency labels indicate when the test run is at or near the fundamental or torsional frequencies of the building. The amplitude and phase alignment is generally good; see Figure 5.8 which expands the time window of 4820 to 4850s. 78

Figure 5.3: Spectrum comparisons in the N-S direction where the fundamental mode of the building is at 1.7 Hz and the fundamental torsion mode is at 2.35 Hz. Blue is spectrum for the Episensor (CI MIK), red is spectrum for a single phone (Samsung Galaxy S4) and the green spectrum is from the 7-phone stack. The modal peaks are clearly visible in all cases. 79

Figure 5.4: Waveform comparisons for two horizontal components between the stack of 7 phone recordings and the Episensor (CI MIK). (A) N-S component. (B) E-W component. The red time series is from the stacked recordings from 7 phones, and the blue is from the MIK recording. Frequency labels indicate when the test run is at or near the fundamental or torsional frequencies of the building. 80

Figure 5.5: Displacement time series comparisons (high-pass filtered at 0.5 Hz). These recordings are extracted from the period of largest-amplitude response to the shaking that occurred in the N-S direction, between 4820 and 4850 s (See Figure 5.2a for reference). (A) The Episensor (CI MIK) compared with a single phone (Samsung Galaxy S4). (B) The Episensor compared with the 7-phone stack. See Figure 5.8 for the accelerations in the same time range. 82

Figure 5.6: Horizontal component recordings after application of a narrow-band filter 1.69 to 1.71 Hz. Energy can be seen on both components. The upper panel is the component oriented roughly N-S, and bottom panel is the component oriented roughly E-W. The phone used here is the same Samsung Galaxy S4 as the one shown in Figure 5.2. 83

Figure 5.7: Result of rotating the horizontal components of the narrow-band filtered signal (Figure 5.6) anticlockwise 1.5 degrees in order to minimize energy on one component. The upper panel shows the axis aligned to N-S after having rotated the components, the lower panel is the axis aligned to E-W. 84

Figure 5.8: Acceleration record comparison for a single phone and the MIK station for time range 4820 s to 4850s (see Figure 5.2a for reference). The red recording is from a single phone (Samsung Galaxy S4) and the blue recording is from the MIK station. The phase alignment is excellent; the peak amplitudes show some deviation from the MIK recording. Also note that the phones and the MIK station are not in the same location; see the main text for their locations. The color version of this figure is available only in the electronic edition. 86

Figure 5.9: Spectrum comparisons for the E-W direction where the fundamental mode of the building is at 1.15 Hz and the fundamental torsion mode is at 2.35 Hz. Blue is spectrum for the Episensor (CI MIK), red is spectrum for a single phone (Samsung Galaxy S4) and the green

spectrum is for the 7-phone stack. The color version of this figure is available only in the electronic edition.....87

Figure 5.10: A partial screenshot of the Millikan Library Cloudlet Display during the shaker tests. The display shows (from left to right, along the top): the instantaneous displacement of each sensor in the E-W direction (in 0.1 mm units); the displacements in the N-S direction; inter-story drift (a dimensionless measure of the difference in displacements between floors divided by floor height); a table of maximum displacements observed on each floor over the previous 10 seconds, 1 hour, and 1 day, for each measured axis. The lower left plot shows the displacement as a function of time for the sensor on the 9th floor. Note the sinusoidal oscillations in the N-S curve in the lower plot (red points) compared to the near-zero displacements in the E-W curve (blue points). The color version of this figure is available only in the electronic edition.....88

List of Tables

Table 2.1: Performance of the ANN algorithm. Performance of classifier when applied to earthquake and non-earthquake data not used to train the ANN algorithm. In the case of earthquake data the percentage of records that were correctly classified as earthquakes is shown along with the number of records (in parentheses) for various earthquakes recorded within various distances of the epicenter. For the every-day human activity data, the percentage correctly identified as non-earthquake and falsely identified as earthquakes is shown. 11

Table 2.2: Accuracy score for ANN classifier with 10-fold cross validation. The score row shows the accuracy score for each run defined as 22

Table 2.3: Simulated network detection performance for US earthquakes. Simulated phone triggers from 2014 La Habra M5 earthquake and 2004 M6 Parkfield earthquake were used to test the network detection algorithm. The magnitude, location and origin time estimates and errors are given for the initial MyShake estimates 24

Table 2.4: Simulated network performance for various phone densities. N is the number of randomly distributed stations within a 1°x1° box (~111x111 km); we did 1000 simulations in each case for a M6.0 earthquake. The location errors are the differences between the true earthquake location and the estimated earthquake location. The origin time errors are the time difference between the true earthquake origin time and that estimated. The detection time is the time after the true earthquake time that the algorithm detects it. In all cases we show the average value ± standard deviation. The last column shows the number of simulations in which the earthquake was not detected. 25

Table 5.1: Smartphones used in the test. Brand and model of the phones are shown, N is the number phones used in the test. Acc. Type is the model of the accelerometer. Max Range and Resolution show the range of the amplitude and the smallest measurable value that the sensor can measure. The phones flagged with * indicate that the phone is used in the 7-phone stack. The resolution values are from the sensor specifications. 77

Acknowledgments

I came to Berkeley Seismology Lab (BSL) with a Civil Engineering background in Aug. 2011. With little earth science background, I immediately learned the huge difference between science and engineering. I really appreciate my great advisor and friend Richard Allen, who gave me patient guidance on research as well as many aspects in my life to overcome the initial frustration. As an advisor, he shows me how to do great research in science, how to find the directions for future research, and most important of all, how to communicate the science to the public. All these training start to turn me from a graduate student into an independent researcher. He has a very open mind with a great vision to allow me to try new things and gives me full support to pursue the things I really like. I remembered when I first started to learn machine learning and tried to apply it in the seismological questions, not so many people in seismology communities appreciate that, and most of them think this is just engineering application and no science behind it. But Richard is different, he encourages me to learn and provides me a lot of resources on campus to push this forward. He also encourages all his students to do internships during the PhD to have a taste of the industry and then to decide whether we like academia or industry after we graduate. Besides research, he is always a good friend and gives me a lot of suggestions for my career choice and personal life. I really enjoyed my discussion with him about anything related to my career, and he always shares his insights and personal experiences to help me moving forward. I feel so lucky that I have a great mentor in my life during my PhD!

My PhD project is slightly different than most of the projects in the traditional sense that I collaborate a lot with the industry, especially I want to thank members from the MyShake team: people from T-Mobile Deutsche Telekom Silicon Valley Innovation Center: Louis Schreier, Roman Baumgaertner, Siddartha Pothapragada, Garner Lee, Arno Puder, Pooja Kanchan, Young-Woo Kwon, and people from BSL: Stephen Allen, Stephen Thompson, Asaf Inbal, Sarina Patel, Jennifer Strauss, Kaylin Rochford, Doug Neuhauser, Stephane Zuzlewski, and Jennifer Taggart. Without the effort from them, MyShake will only exist on the paper instead of the whole global seismic network.

I would like to thank the other two members of my dissertation committee, Professor Douglas Dreger and Professor Jonathan Bray – both of whom provided valuable feedback not only on the material in this thesis, but also as members of my qualifying exam committee and in their courses that I've had the pleasure of taking. And I would like to thank another member of my qualifying exam committee – Professor David Romps also for his great help and discussion.

I would specially thank Roland Bürgmann, Douglas Dreger, and Barbara Romanowicz to allow me to come to their group meetings. I expand my knowledge by learning from others and discussing various topics in these group meetings.

I also want to thank all the professors in EPS who gave me many insightful views on various topics, such as James Bishop, William Boos, Bruce Buffett, Kurt Cuffey, Inez Fung, Raymond Jeanloz, Michael Manga, Mark Richards, David Shuster, Nicholas Swanson-Hysell, Chi-Yuen Wang. Every time when I talk with them, I always can get something useful that completely out of my mind.

A big thanks to the previous and current members in the earthquake early warning research team: Christine Ruhl, Angie Chung, Asaf Inbal, Ivan Henson, Sarina Patel, Jennifer Strauss, Diego Melgar, Ronni Grapenthin, Serdar Kuyuk, Lingsen Meng, Holly Brown, Ran Novitsky Nof, Simona Colombelli. You all helped me to improve my skills in the study of earthquake early warning systems.

My officemates Felipe Orellana, William Hawley, Chris Johnson, Mong-Han Huang, Kathryn Materna, Runze Miao are part of my life almost everyday during the PhD study. I cannot forget all the beers we drank and all the fun we had in the office! Thanks all of you.

Also, thank all the previous or current members in BSL and EPS department: Sanne Cottar, Aurelie Guilhem, Rob Porritt, Shan Dou, Allen Zhao, Jiajun Chong, Yang Zhao, Scott French, Andrea Chiang, Sierra Boyd, Shuai Zhang, Heidi Fuqua, Eloisa Zepeda, Avinash Nayak, Robert Martin-Short, Alex Robson, Nate Lindsey, Li-Wei Chen, Sevan Adourian, Yuexin Li, Nam Maneerat, Robert Nadeau, Peggy Hellweg, Horst Rademacher, Taka'aki Taira, Fabia Terra, Pete Lombard, Charley Paffenbarger, Dan Frost, Satish Maurya, Artie Rodgers, Baptiste Rousset, Corinna Roy, Lian Xue, Kang Wang, Margie Winn, Kathryn Wooddell, Brent Delbridge, Seth Saltiel, Ian Rose, Zack Geballe, Amanda Thomas, Kelly Wiseman, Noah Randolph-Flagg, Jiabin Liu, Xueling Liu, Sean Wahl, etc. I really appreciate that you accompany with me in my life at Berkeley, and form a precious friendship.

I also want to thank my best friend Cheng Cheng who shares a lot of happiness, frustration with me during my stay at Berkeley. He is like my brother over these years, and makes the PhD life easier for me as well.

Of course, I owe too much for my family. My parents Ping Gong and Xianzhong Kong, they offered me all their support in my life without asking for any return. Without them, I won't be here today. My parents-in-law Xiaozhen Wu and Zhiping Lu, who make my PhD life easier by helping me to look after my babies. My wife Fan Lu, who was my classmate when we were back in China, and she is my soul mate that makes me more complete in every possible way. My daughter Fanyu Kong and son Fanqi Kong were born during my PhD, they make my life so much richer that I hope they will appreciate in the future that their father is brave enough to take them to the world during the busy PhD study.

Finally, I thank all the people who appeared and will appear in my life, and you all make my life experience unique and wonderful! I am extremely lucky that I have so many great people appear during my journey of life, it is a wonderful life!

Chapter 1 Introduction

In the last century, we witnessed so many devastating earthquakes, to name a few, the 1906 Great San Francisco earthquake, 1976 M7.5 Tangshan earthquake, the 2004 M9.1 Indonesia earthquake, 2011 M9.0 Tohoku-Oki earthquake, and so on. Seismologists are constantly trying to develop ways to inform and to protect people from earthquake hazards. The deployment of global seismic network, greatly improved our ability to monitor and study the earthquakes. We started to have better understanding of how and where do the earthquakes usually occur. This thesis summarizes the development and progress on building a global smartphone seismic network by taking the power of crowdsourcing. The approach taken here is a combination of data science with the earth science. The following chapters give the detail of how did we build this network with the initial observations.

In Chapter 2, this thesis talks about the motivation behind this project and the initial investigation of building this global smartphone seismic network. Before we build this smartphone seismic network, we asked a few questions: (1) can the sensors inside the smartphones sensitive enough to record earthquakes? We report the noise floor tests we did to answer this question. We found most sensors inside the smartphones could detect earthquakes larger than M5 within 10 km with some best sensors starting to see M3.5 earthquakes in certain frequency bands. A trend of decreasing the noise level on newer phones has also been observed to suggest that the sensors may get better in the future to record smaller earthquakes. (2) Can the smartphones reproduce the shaking of the earthquake well even when they are not fully coupled with the ground? The answer confirmed by doing shake table tests with putting the phones on the table directly. Even with large shakings, the phones could record the shaking reasonably good, especially for the frequency content. (3) Can we recognize earthquake movements recorded on the phone from the daily human activities? To address this core problem, we build a prototype system to collect the human activity data for 4 months and the shake table tests data as well as simulated earthquake data. With the data we collected, an Artificial Neural Network (ANN) approach is designed to recognize earthquake-like movement from the daily recordings on the phones. Different features that characterize the difference between earthquake-like and human-like motions are compared and selected carefully to feed into the ANN. After training, the ANN could recognize most of the earthquakes within 10 km above M5.0, and commit about 7% false positive rate. In the end of this chapter, we also talk about a simple network detection algorithm to aggregate results from multiple phones in a region. 1000 simulations were conducted to validate the performance of the network detection algorithm.

After Chapter 2, MyShake was release to the public on Feb 12th 2016. Within a short time, MyShake users occupied the 6 continents and started to contribute data to the central server. Chapter 3 shows the initial observations from the data collected from this network within 6 months after its release to the public. Earthquakes from M2.5 to M7.8 with useful waveforms were recorded by the MyShake user globally. One particular earthquake event – the M5.2 Borrego Springs earthquake occurred on Jun 10th 2016 is showing here. Most of the phones triggered either on the P wave or the S wave depending on the distances from the event, the site effect and the noise level of the sensors. The record section plot of the waveforms recorded by MyShake users shows clear moveout of the P and S waves at difference distances. The Peak

Ground Accelerations from the phones show a similar attenuation relationship as that from the seismic network, even though the amplitudes from MyShake is higher, may due to the local site effect, and the fact that the phones are in buildings and put freely on desks and so on. Besides, you can see the comparison of the MyShake recordings with a nearby seismic station at about 40 and 100 km for the M5.2 earthquake. At closer distance, the matching of the waveforms are good, and even the P wave matches well. But at about 100 km, due to the high noise level on the phones, the P wave portion is buried into the noise but you can still see the S wave and the strongest portion matching well. This will be what we expect to see for the phones close and far away from the earthquakes. Finally, in this chapter, we will also show the observations of waveforms recorded by MyShake have clear P wave arrivals, especially for larger magnitude earthquakes, even at farther distances, we can still see it very clear. One example is from M7.8 Ecuador earthquake recorded by one MyShake user at about 170 km, we can clearly see the P wave. These observations are promising both in terms of real-time earthquake early warning and other seismological applications.

Chapter 4 continues to show the application we can do using the waveforms recorded by MyShake users. We start to compute the most important earthquake parameters, such as location, origin time, and magnitude. We first select the earthquakes that have more than 5 useful earthquake waveforms recorded by MyShake. We manually picked the P and S wave arrivals from each waveform. With 18 events with both P and S phases, we could estimate the locations of the earthquake with a mean error of 8.4 km and standard deviation 7.6 km relative to USGS locations. Since the high noise level of some of the phones, not all the waveforms can pick the P wave arrivals. Many of the waveforms can only pick the S wave arrivals, depending on the magnitudes of the earthquake and epicentral distances of the phones. We could use only the S wave arrivals from each earthquake to estimate the location and origin time of the earthquake using grid-search approach by assuming the S wave speed as 3.55 km/s. The location estimations using the S wave arrival times have a mean of 17.8 km with standard deviation 22.8 km in terms of error from the USGS catalog locations. The origin time estimation residuals with the USGS catalog origin time have a mean of 0.1 s with standard deviation 6.9 s. Generally, we found that with a better coverage of the phones, we could do a better job in terms of estimating the location. Besides the locations and origin times of the earthquakes, we could also estimate the magnitudes from the waveforms. Most of the errors of the estimated magnitudes with the catalog ones are within 1 magnitude unit with the mean difference 0.3 and standard deviation 0.4 (estimated – catalog magnitude). As MyShake recordings are generally larger amplitude than the nearby seismic stations, we also tried to find the optimum factor to scale down the MyShake recording amplitudes to reduce this effect by minimizing the magnitude estimations compare with USGS magnitudes. We found the factor 1.6 is the best for the current dataset, and with this correction, the mean magnitude difference between MyShake estimated magnitudes and the USGS ones reduced from 0.3 to 0.1. This chapter illustrates the usefulness of the waveforms recorded by MyShake users, which could be used to provide valuable information about earthquakes for places where no or few seismic stations exists.

In addition to the seismological applications we show in the previous chapters, chapter 5 also explores using the sensors inside the smartphones to extract the fundamental frequency of the buildings as a way to monitor the health state of the buildings. In order to prove the sensors inside the phones could be used as this purpose, we conducted the shaker test to shake a 9-story building on Caltech campus. Testing phones were placed on the floor of the top story and recorded the shaking during the test. We show we could extract the fundamental frequencies

from the recorded waveform from single phone by transforming it to the frequency domain. If we have multiple phones nearby, we could stack the waveforms to lower the noise level and improve the signal noise ratio to see smaller amplitude peaks. As we compare the displacement derived from the phone recording by double integration to that from the reference sensor, both phase and amplitude match well. The test is similar to the results expected for stationary phones resting on stands at night, or when placed on a desk or left in a bag on the floor. The results here illustrate the potential of using MyShake-enabled personal smartphones to record building shaking resulting from nearby earthquakes and using that data to extract the building characteristics. In this chapter, we also present a method to determine the orientation of the smartphone if its orientation is not known, but prior information about the building characteristics is available.

Chapter 2 MyShake methodology

Published as: Qingkai Kong, Richard M Allen, Louis Schreier, and Young-Woo Kwon. 2016. "MyShake: A Smartphone Seismic Network for Earthquake Early Warning and beyond", Science Advances 2 (2): e1501055–e1501055. doi:10.1126/sciadv.1501055.

2.1 Abstract

Large magnitude earthquakes in urban environments continue to kill and injure thousands to hundreds of thousands of people inflicting lasting societal and economic disasters. Earthquake early warning (EEW) provides seconds to minutes of warning allowing people to move to safe zones and automated slowdown and shutdown of transit and other machinery. The handful of EEW systems operating around the world use traditional seismic and geodetic networks that only exist in a few nations. Smartphones are much more prevalent than traditional networks and contain accelerometers that can also be used to detect earthquakes. In this chapter, we report on the development of this new type of seismic system--MyShake--that harnesses personal/private smartphone sensors to collect data and analyze earthquakes. We show that smartphones can record magnitude 5 earthquakes at distances of 10 km or less, and develop an on-phone detection capability to separate earthquakes from other every-day shakes. Our proof-of-concept system then collects earthquake data at a central site where a network detection algorithm confirms that an earthquake is underway and estimates the location and magnitude in real-time. This information can then be used to issue an alert of forthcoming ground shaking. MyShake could be used to enhance EEW in regions with traditional networks, and could provide the only EEW capability in regions without. In addition, the seismic waveforms recorded could be used to deliver rapid microseism maps, study impacts on buildings and possibly image shallow earth structure and earthquake rupture kinematics.

2.2 Introduction

Large magnitude earthquakes in densely populated regions do not occur very frequently, but they can kill tens to hundreds of thousands of people, injure many more, and cause substantial financial loss [Holzer and Savage, 2013]. Earthquake Early Warning (EEW) systems can detect the location and magnitude of an earthquake in a few seconds, and issue a warning to the target area before the damaging waves arrive [HEATON, 1985; Allen and Kanamori, 2003]. This new technology can reduce the fatalities, injuries, and damage caused by earthquake by alerting people to take cover, slowing and stopping trains, opening elevator doors, and many other applications [Allen, 2011]. The development of EEW to date has largely focused on the use of traditional seismic and geodetic networks, which only exists in a handful of countries around the world [*Allen et al.*, 2009b]. Smartphones are much more prevalent, and have a variety of built-in sensors and communications. There were 2.6 billion smartphones worldwide in 2014, and this number is expected to pass 6 billion by 2020. In this chapter, we report on development of MyShake, a crowdsourcing project [*Allen*, 2012b] to harness the accelerometers in personal smartphones to record earthquake-shaking data for research, hazard information and earthquake early warning.

We build on other crowdsourcing projects in seismology. The Quake Catcher Network (QCN) and Community Seismic Network (CSN) primarily use low cost MEMS accelerometers that plug into computers and can be installed in buildings to detect earthquakes [*Cochran et al.*, 2009b; *Clayton et al.*, 2011]. These networks consist of a few hundred to a few thousand accelerometers, but are limited by the need to pass hardware from the network operators to the users. By using the sensors in smartphones, we only need to pass software from the network operators to users, which is relatively simple using the Google Play and iTunes store. The CSN also explored the use of smartphone accelerometers. Their approach was to ask if newly incoming data is similar to previously defined human activities. If not, it is treated as an anomaly, and communicated to a processing center where a picking algorithm will determine if it is earthquake or not [*Faulkner et al.*, 2011; *Olson et al.*, 2011]. Our MyShake design is different in that we use past earthquake information to develop a classifier algorithm to identify earthquake shaking on a single phone, and then communicate with a centralized processing center (CPC). Previous work has also demonstrated that the GPS sensors on smartphones (rather than the accelerometer) can be used to detect earthquakes and potentially provide a warning [*Minson et al.*, 2015]. To date, this has been shown to be possible on dedicated smartphones, but not on personal smartphones. Another crowd-sourcing project is using twitter to detect earthquakes. A tweet-frequency time series constructed from tweets containing the word earthquake in various languages, and an algorithm is used to identify possible earthquakes [*Earle et al.*, 2011]. Finally, the USGS Did You Feel It (DYFI) system is a web-based approach for collecting reports of shaking and damage as experienced by individuals. The reports are converted into intensity and used to generate detailed shaking intensity maps when enough people report [*Survey and Dewey*, 2005; *Wald et al.*, 2011]. The intensity estimate relies on subjective descriptions by the reporter. By using smartphone sensors, MyShake utilizes the power of crowd sourcing, while also reporting shaking timeseries and accurate locations.

The MyShake network builds on some initial work at UC Berkeley to determine the quality of the accelerometers in smartphones [*Dashti et al.*, 2011]. We have extended this work to develop an android-based application that runs efficiently on the users' smartphone and detects whether the movement of a phone is likely caused by an earthquake verses other human activities. It sends this information back to our processing center where a network detection algorithm confirms that an earthquake is underway. The location, origin time, and magnitude of the earthquake are then determined based on multiple triggers from the network of phones. This information can be used to estimate the shaking intensity and remaining time until damaging waves arrive at a target location. In the following we detail (a) size and proximity requirements for earthquake signals to be recorded by smartphones, (b) development of our on-phone detection capability to distinguish earthquakes from other shakes, and (c) design of a network detection algorithm to operate at the processing center to confirm when an earthquake is underway, locate and characterize it. This has been achieved within the real-world constrains of building an android application that runs in the background on private phones without draining power.

2.3 Results

To better understand which earthquakes we can record on smartphones, we determined the noise floor [*McNamara and Buland*, 2004] of the accelerometers on multiple android phones by placing them in a basement and allowing them to record for one month. The noise floor of the phones contains the internal noise of the phone itself plus other environmental sources in a quiet basement. Once this level is known, we can assess the necessary size of earthquakes such that the ground shaking amplitude exceeds the noise. Figure 2.1 compares the noise floor of the test phones to the amplitude of shaking for various magnitude earthquakes at 10 km [*Clinton and Heaton*, 2002]. All phones are sensitive to the shaking for M5 or larger earthquakes 10 km or less from the phone in the frequency range of 1 to 10 Hz, and they are capable of recording the longer periods of larger magnitude events. There is a gradual improvement in the sensor capabilities with the release date of the phone (see the color change from cold to warm). The more recent phone models are sensitive to shaking for M3.5 at 10 Hz. The in-phone accelerometers can record shaking for the earthquakes that do damage in the frequency range that causes most damage (1 to 10 Hz). Also, we expect the quality of the sensors in phones to improve further with time. The HP MEMS accelerometer (blue, Figure 2.1 was recently developed for seismic imaging applications [*Homeijer et al.*, 2011]. It is currently too expensive for inclusion in smartphones, but illustrates that MEMS sensors can have similar capabilities to more traditional strong motion sensors (station BKS, Figure 2.1).

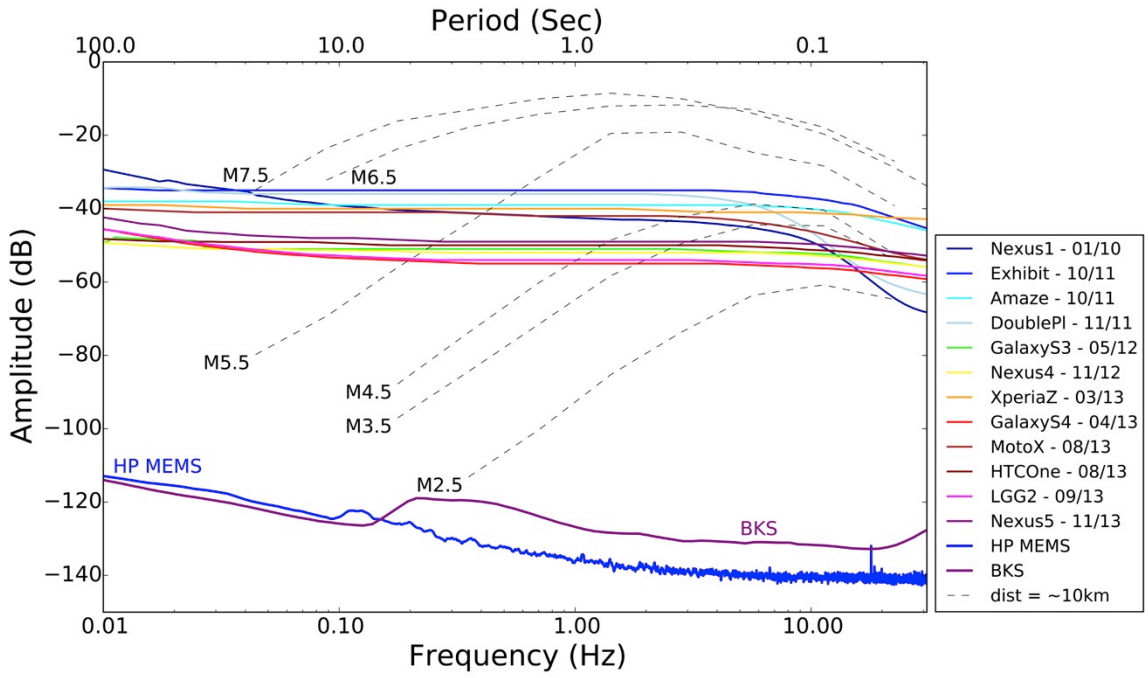


Figure 2.1 Noise floor of the phones. Noise floors of the smartphones color coded by the phone release date (also shown in the legend as MM/YY). Dashed black lines are typical ground motion amplitudes of earthquakes 10 km from the epicenter for various magnitudes. Noise floor for high quality MEMS sensor (HP MEMS - blue) and a typical force-balance accelerometer from a regional network (BKS in northern CA - purple) are also shown.

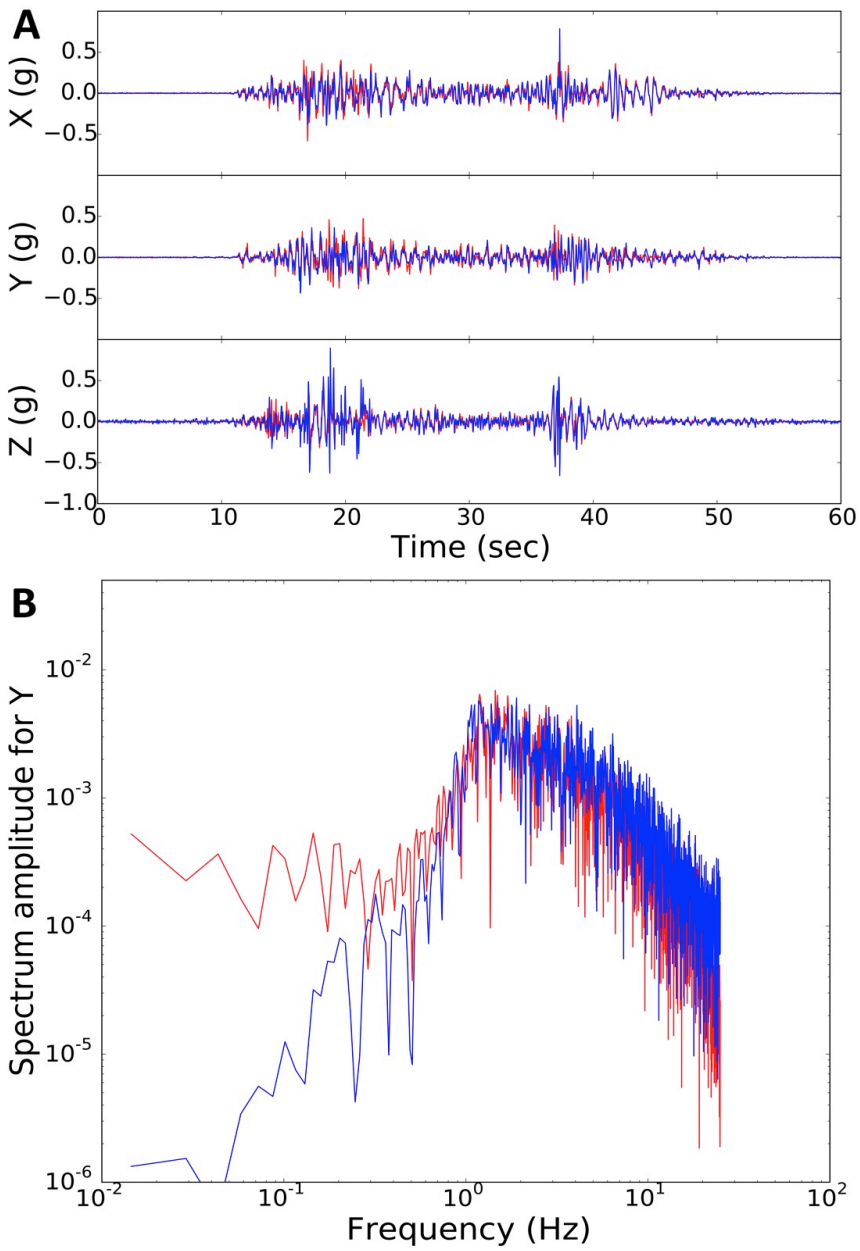


Figure 2.2: Three-dimensional shake table test. The input seismogram is from a real earthquake that has been modified for IEEE-693-2005 tests. (A) Waveform comparison between phone (blue) and reference accelerometer (red) recordings from an input signal that has peak acceleration of 0.5g. (B) Spectrum comparison of Y components. The X and Y components are in the plane of the phone, which is lying flat on the horizontal shake table and is not attached. The Z component is perpendicular to the plane of the phone and is vertical for this test.

Next we must determine how well phones can record the true shaking in an earthquake. Both the quality of the sensor and how well coupled the phone is to the ground play key roles here. We deployed multiple phones on shake tables to answer this question; some were bolted to

the table, others could freely slide. Our results confirm previous work [*D'Alessandro and D'Anna*, 2013] that phones bolted to shake tables are capable of recording ground motion accurately between 0.5 and 10 Hz. We also tested phones placed freely on the shake table, since personal phones are not bolted to the ground. Figure 2.2 shows a three-dimensional shake table test with the peak acceleration 0.5g. The phone under testing had some relative motion with the table, but minimal. We can see that the waveform of the phone and the reference accelerometer are very similar, and the frequency response of the phone acceleration is good from 0.5 Hz up to 10 Hz. In a one-dimensional shake table test with a sweep signal (increasing amplitude and frequency gradually), we found that it was not until the horizontal accelerations reached certain threshold, in this case ~0.3g and above ~3Hz, that we started to see sliding. When the phones slid, it had the effect of clipping the peak amplitudes but the frequency content remained similar (Figure 2.3). This is a limitation of the data recorded, and we must recognize that recorded amplitudes are lower bounds on the actual value.

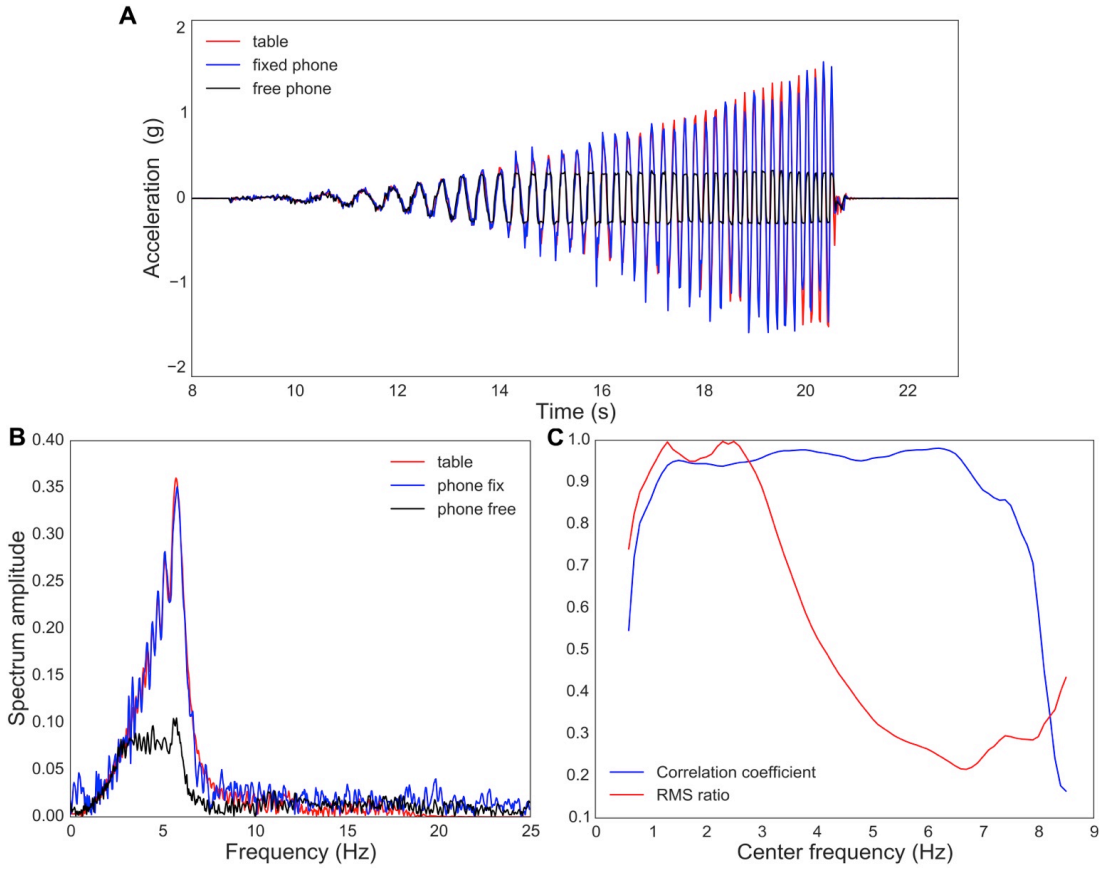


Figure 2.3: Shake table test with an input sweep signal (0.5-7Hz). (A) Waveform comparison between a phone fixed on the table (blue), a phone placed freely on the table (black) and the reference accelerometer attached to the table (red). (B) Frequency domain comparison of the signals in (A). (C) Calculated correlation coefficient and RMS (Root Mean Square) ratio between the signal recorded by the phone placed freely on the shake table and the reference accelerometer. The correlation coefficient is a measure of the phase match and RMS is a measure for amplitudes match. We use 1 Hz frequency band to filter the record and calculate the

coefficient with a step frequency 0.1 Hz. The x-axis is the center frequency of the frequency band. The correlation coefficient shows how well the phase is recorded by the phone, and the RMS ratio shows the amplitude recovery. Above 2-3 Hz the phone starts to slide so the full amplitude is not recovered, however, the phase is recovered up to 7-8 Hz.

Given that a smartphone can record earthquake shaking, the key challenge for a smartphone network using private/personal phones is being able to separate earthquake shaking from every-day motion of the phone. Figure 2.4a shows 12-hours of 3-component acceleration data that was recorded on a smartphone. It contains both human activities and the M6.0 Napa earthquake on August 24, 2014 [*Brocher et al.*, 2015] at the very end of the waveform. Figure 2.4b shows the zoomed in view of the accelerations associated with the Napa quake recorded on the same phone.

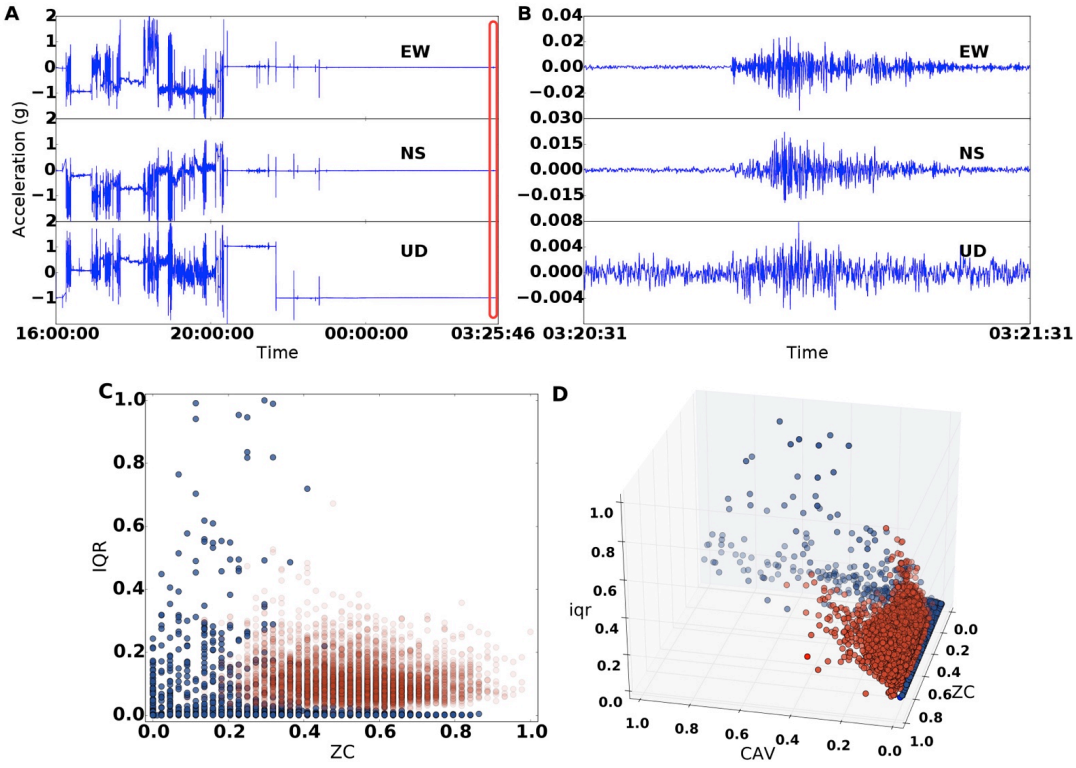


Figure 2.4: Earthquake recorded by phone and classifying earthquakes. (A) Example 12-hour 3-component acceleration record from a private/personal Samsung Galaxy S4 phone starting at 4 pm August 23, 2014. It shows the accelerations of every-day human motions for the first ~8 hours, then appears stationary during the night. The red box at the end of the figure highlights the time window of figure b. (B) 1 minute of data from the period shown in (A) at the time of the M6 Napa earthquake 38 km from the phone. The earthquake occurred at 3:20:44 am local time. (C) Scaled feature plot showing IQR versus ZC for the classifier training dataset. The blue dots are the centroids of human activities, and the red dots are the earthquake features. (D) 3D plot of the 3 features we used to distinguish earthquakes. Adding the CAV to IQR and ZC drags some of the human activates (blue dots) to the third dimension but not the earthquake data, this helps improve the results.

In order to develop an algorithm to separate earthquake shaking from human activities, we first developed an application for android smartphones to trigger on significant motions, and send the data to a central processing center (CPC). It has been designed for distribution to personal/private phones and has a trigger algorithm that runs in the background monitoring the accelerometer continuously. It uploads parameters and data to our CPC when triggered. Functionality at the CPC allow us (a) to monitor and change the operational parameters on the user phones, (b) collect heart-beat and state-of-health information from the phones, (c) collect autonomous phone-trigger information, (d) trigger phones from CPC to record data, and (e) upload waveform data for autonomous and CPC triggers. A small release of MyShake in November 2014 deployed the application on 75 phones. A key issue for a crowdsourcing application to be successful is minimizing the impact on the users: in the case of a phone this means minimizing power usage. The MyShake application currently uses about the same power that a smartphone uses when it is on, but is not being used. For most users, a phone running MyShake does not need to be charged more than once every 24 hours.

Using the data collected, we have developed an Artificial Neural Network (ANN) approach to identify the different characteristics of earthquake and human motions (see supplementary material for details). The algorithm assesses 2 sec windows of data and determines if the motion is likely an earthquake or not. We must first train our algorithm. The training data comes from three sources: every-day motion recordings uploaded to our CPC from the MyShake release as described above, phone recordings of earthquakes from shake table tests, and seismic data from traditional networks in Japan that was modified to reproduce smartphone-quality records, which is described in supplementary material. We tested a total of 18 characteristics identifying the 3 best features: Interquartile range of the acceleration vector sum (IQR), the maximum zero crossing rate (ZC), and the cumulative absolute velocity of the acceleration vector sum (CAV). IQR is an amplitude parameter that shows the middle 50% range of amplitude of the movement. ZC is a simple frequency measure that counts the number of times when the signal crosses baseline zero. CAV is a cumulative measure of amplitude on the three components in the time window and is determined as follows:

$$CAV = \int_0^2 |a(t)|dt \tag{2.1}$$

Where a(t) is vector sum of the 3 components acceleration.

Figure 2.4c shows how IQR (a measure of amplitude) and ZC (a measure of frequency) separate earthquakes from non-earthquake motions. Earthquakes are high frequency with moderate amplitudes while every-day motions are lower frequencies but high amplitudes or very low amplitudes but high frequencies. The IQR and ZC are the best two parameters to separate earthquakes, but adding CAV can provide some additional information to help improve performance (Figure 2.4d).

The trained ANN algorithm is then applied to US earthquake data modified to phone-quality records and a separate set of every-day motion data (Table 2.1). 98% of the earthquake records within 10 km of the events are recognized as earthquakes; the success rate reduces with increasing distance and decreasing magnitude as expected. 93% of the every-day motions are correctly recognized, meaning that for an operational system we should expect ~7% of phone triggers to be false (earthquake) triggers.

Table 2.1: Performance of the ANN algorithm. Performance of classifier when applied to earthquake and non-earthquake data not used to train the ANN algorithm. In the case of earthquake data the percentage of records that were correctly classified as earthquakes is shown along with the number of records (in parentheses) for various earthquakes recorded within various distances of the epicenter. For the every-day human activity data, the percentage correctly identified as non-earthquake and falsely identified as earthquakes is shown.

Earthquake classification	Within 10 km	Within 20 km	Within 30 km
1989 Loma Prieta M7	100% (2/2)	100% (4/4)	100% (11/11)
1994 Northridge M6.7	100% (4/4)	100% (19/19)	100% (29/29)
2004 Parkfield M6	95% (19/20)	90% (35/39)	86% (36/42)
2014 Napa M6	100% (2/2)	75% (6/8)	42% (10/24)
2014 La Habra M5.1	100% (13/13)	42% (22/52)	25% (30/120)
Human activity classification	non-earthquake (correct)	earthquake (false)	
20150201-20150228	93% (3562/3823)	7% (261/3823)	

The final component of our system is a network detection algorithm running at the CPC to confirm when an earthquake is underway, and estimate source parameters from multiple triggered phones in a region. When a phone determines that it is recording an earthquake, two types of data are passed to the CPC: (a) the trigger information including trigger time, phone location, and the maximum amplitude of the 3 components, and (b) waveform data that contains 3-component acceleration from 1 minute before the trigger to 4 minutes after. The trigger information is easier to upload rapidly via cellular or Wi-Fi networks and is what we use for real-time processing. The waveform data is currently uploaded with a lower priority and only uploaded when the phones are connected to Wi-Fi and power.

Our first-generation network detection algorithm is based on current earthquake early warning ElarmS-2 methodologies [*Kuyuk et al.*, 2014]. It searches for a temporal and spatial cluster of triggers, and requires greater than 60% of operating active phones to have triggered within a 10 km radius region for an event to be declared, see the supplementary material for details. Once an event is created, the algorithm will continue to update the origin time, location, and magnitude of the earthquake based on the continual flow of trigger information. Currently, the origin time is set to the earliest trigger time, and the centroid of the all the triggered phones within 10 km of the phone trigger is used as the epicenter. Our first generation magnitude estimation is based on expected ground shaking amplitude as a function of distance. We use the Peak Ground Acceleration (PGA) and the distance of the station to estimate the magnitude using the following regression relation based on the earthquake data from Japan that was modified to reproduce smartphone-quality records:

$$M_{est} = 1.352 \times \log(PGA) + 1.658 \times \log(distance) + 4.858 \qquad (2.2)$$

Where PGA is the maximum absolute amplitude from the 3-component acceleration, and distance is the epicentral distance derived from the phone location and estimated location of the earthquake. Figure 2.5 compares estimated magnitude and the real magnitude for both individual phone (blue dots) and the average event estimates (red pluses). We can see that, most of the estimated magnitudes are within 1 magnitude unit for individual phone, and all average event estimates are within 1 magnitude unit. When the network consists of many more phones, we might expect the uncertainty in the magnitude to be reduced. However, we must also recognize that phone-based amplitude estimates must be treated as lower bounds given the possibility of decoupling. Given these uncertainties, it is clear that having even a single observation from a traditional seismic station could make a significant difference providing some "ground truth" to the magnitude estimate.

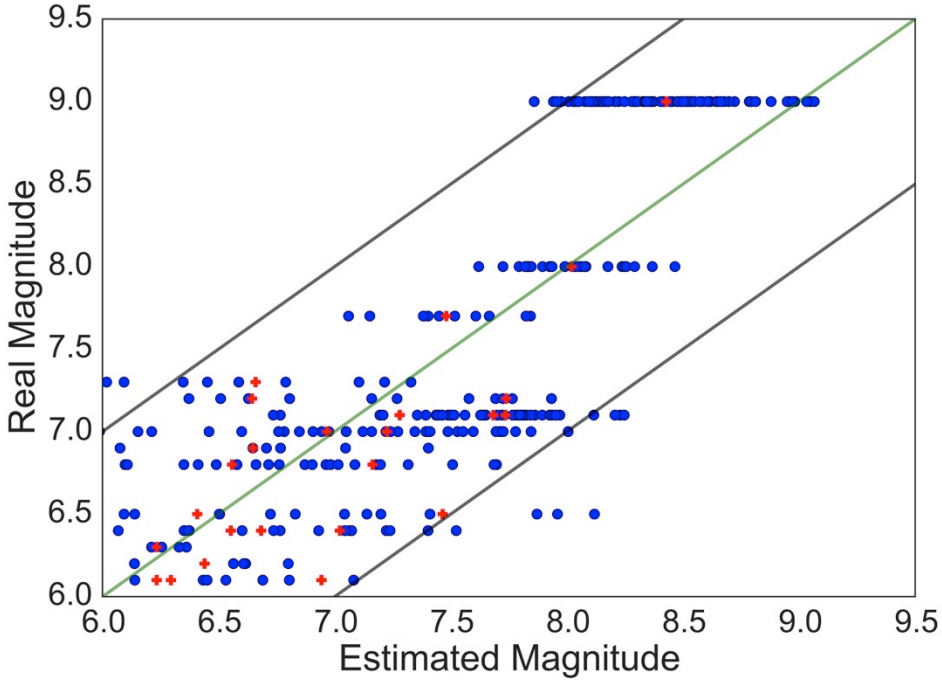


Figure 2.5: Estimated magnitude. Comparison of our estimated magnitudes with the real magnitude for earthquakes in Japan using phone-like data. The green line is the 1:1 line, and the two grey lines are the 1 magnitude unit shift, each blue point is the magnitude estimate at a single simulated phone. The red pluses are the average event estimates, which is the average of multiple single phone estimates.

The final step for an alert is to estimate the shaking intensity and time till shaking at a users' target location. This is relatively straight forward using the estimated event epicenter, origin time and magnitude, the users' location, and S-wave traveltime curves and ground motion prediction equations [*Worden et al.*, 2012] just as with the current EEW system in California.

It is a known problem that magnitude estimates based on peak shaking observations from seismic stations saturate [*Anon*, 2011; *Colombelli et al.*, 2012]. This will also be a problem for

MyShake. There are several possible improvements. Firstly, the smartphone-based magnitude estimate could be improved by updating the magnitude based on the area experiencing strong shaking. Stronger magnitude earthquakes cause strong shaking over large areas. Another possibility is to make use of GPS-based permanent ground displacements as is being done with the more traditional network-based early warning systems [*Colombelli et al.*, 2013; *Grapenthin et al.*, 2014]. It was recently shown that smartphone-based GPS observations could be used for EEW [*Minson et al.*, 2015]. The challenge when using only GPS on smartphones is that GPS is very power-hungry. A possible hybrid, would be to start monitoring the GPS on a phone when the MyShake classifier identifies an earthquake. This could provide an updated magnitude estimate that does not saturate and would not suffer from the power issues associated with an only-GPS approach.

We applied the network detection algorithm in a simulated real-time manner to phone-like triggers for US earthquakes (Table 2.1). For the stations that are close to the epicenter (within 10km), almost all stations are triggered. Figure 2.6 shows performance snapshots for the M5.1 La Habra earthquake [*Donnellan et al.*, 2015], which had the poorest success rate in triggering on individual phone-like waveforms due to the relatively small magnitude compared with other test earthquakes (Table 2.1). The figure shows the location of the triggers at each time step; the radiating nature of the ground motion and associated triggers is clearly seen. The earthquake is first identified 5 sec after the origin time (Figure 2.6b). The error in the initial magnitude estimate is 0.1 magnitude units, the location error is 3.8 km, and the origin time error is 1.7 sec (Table 2.3). The performance of this MyShake simulation is similar to the actual performance of the real-time ShakeAlert/ElarmS EEW system, which issued its first alert 5.3 seconds after the origin time with an initial magnitude error of 0.8, location error of 1.5km and origin time error of 0.2 sec. In reality when we have a denser phone network, we would expect to detect the earthquake faster. Movie S1 and S2 show performance animations for the 2014 La Habra and 2004 Parkfield events [*Langbein et al.*, 2005] respectively.

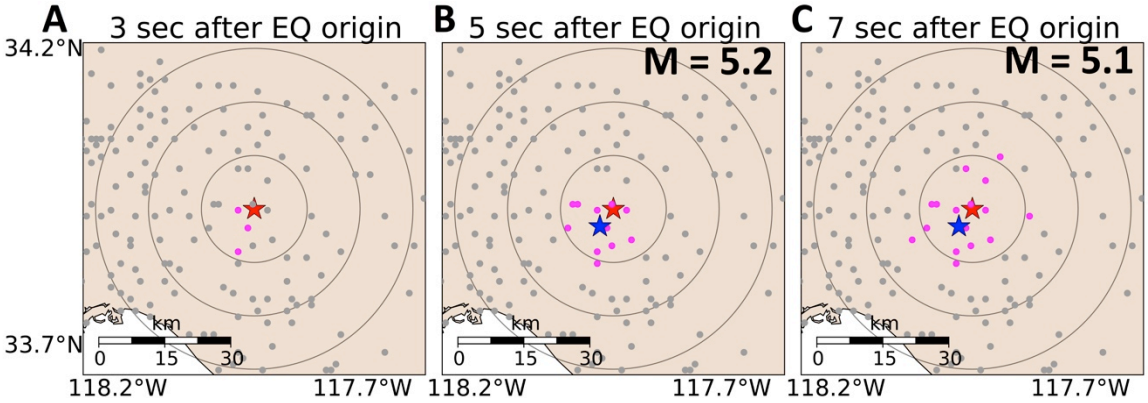


Figure 2.6: Snapshots of trigger detections for the 2014 M5.1 La Habra earthquake simulation at 3, 5 and 7 sec after the event origin time. Grey dots are stations; pink indicates a trigger. The true earthquake location is the red star with circles at 10, 20 and 30 km radius. The blue star represents the estimated event location, first detected at 5 sec. The magnitude estimate at each point in time is shown upper right.

We also conducted 1000 simulations that incorporate random human-activity triggers in

addition to earthquake triggers to explore system performance for different densities of phones (see supplementary material). We found good performance (similar to the La Habra example) when there are 300 or more phones in a 111 by 111 km region, corresponding to an average distance between phones of 6.4km (Table 2.4). If the number of phones drops to 200 in the same region, then out of 1000 simulations, we found 32 events that were not detected, i.e. 3% of events were missed. In addition to missing some earthquakes, the accuracy of the locations and origin times is degraded. We also conducted a second group of 1000 simulations without earthquakes, just false triggers. None of these generated a false event. This is because we require >60% of active phones within a 10 km radius region to trigger for an event declaration. Our ultimate design goal is to have much smaller distances between active phones than 6.4 km, yet we must recognize that the network algorithm will need to be modified to reflect the active network and these changes may need to happen in real-time.

2.4 Materials and Methods

2.4.1 Data collection

The MyShake application was used to collect all smartphone data used in this study. It can be installed on android phones to record acceleration data. For the noise floor tests and shake table tests, MyShake recorded continuously, and saved the data locally on the phone. The human activity data was recorded using a trigger-based method, 5 minutes of data was collected when the phone satisfied the trigger we described above (see more details in supplementary material at the end of the chapter).

2.4.2 Noise floor test

The noise floor tests were performed by putting smartphones in a quiet basement on the Berkeley campus. The phones recorded continuously for one month at 50 samples per second. The method used to calculate the noise floor (Figure 2.1) is described in [*McNamara and Buland*, 2004].

2.4.3 Shake table test

The shake table tests were conducted at the Pacific Earthquake Engineering Research Center. The phones recorded at 50 samples per second continuously when the shake table simulated the earthquakes. A high quality reference accelerometer also installed on the shake table provided the reference traces. We then compared the recordings from the phones and reference accelerometer both in the time and frequency domain. During the tests, some phones were bolted on the shake table while the rest were placed freely on the table.

2.4.4 Single phone detection algorithm design

The earthquake detection algorithm running on the phones was designed using past earthquake data (from traditional seismic networks but modified to be phone-like quality), shake table data, and human activity data recorded on the smartphones. We used an artificial neural network to design the algorithm to distinguish earthquakes from human activities. The processing of data and steps are described in detail in supplementary material at the end of the chapter.

2.4.5 Network detection algorithm design:

The network detection algorithm is designed based on the current ElarmS-2 methodologies [*Kuyuk et al.*, 2014]. It searches for temporal and spatial clusters of triggers from active phones. We tested the algorithm both on the simulated earthquake data and simulated trigger data (see the chapter supplementary material for details).

2.5 Discussion

The MyShake project to date demonstrates proof-of-concept for a smartphone-based seismic network that provides instrumental recordings of ground shaking in damaging earthquakes, and potentially delivering earthquake early warning. What is key to this study is that the system has been designed for and tested on privately owned smartphones, of which there are billions. To harness the full potential of crowd sourcing, scientists must use sensors that are already being purchased by consumers, we must develop systems that can harness the data from these sensors with minimal impact to the owners, and we must provide the owners with real benefits to participating. MyShake uses the accelerometers on common smartphones, the application is freely available from the Google Play store for easy installation and automatic update, it uses minimal power meaning phones only need to be recharged daily as is common practice, and participation leads to delivery of earthquake hazard information and could include the delivery of earthquake shaking alerts. MyShake details and updates can be found at MyShake webpage.

In the future, existing earthquake early warning systems that use traditional seismic and geodetic networks could benefit from MyShake just as MyShake could benefit from integration of data from traditional networks. As described above, observations from even one traditional seismic station could help reduce uncertainties in MyShake earthquake estimates. Likewise, a handful MyShake phone triggers could be used to confirm a preliminary earthquake detection from one or two traditional network station triggers; most traditional EEW systems require several stations to trigger before issuing an alert. Finally, and perhaps most importantly, MyShake could deliver alerts in regions that have little in the way of traditional seismic networks. This includes Haiti and Nepal that both had recent devastating earthquakes, and other high hazard regions like Iran, Afghanistan, Pakistan, Mongolia, Malaysia, Indonesia and the Philippines. As an example, the recent earthquakes in Nepal are estimated to have killed over 8000, while there are only a handful of seismic stations in the region. Yet, there are an estimated 6 million smartphones in Nepal. Based on the 80 km separation of the M7.8 epicenter from Katmandu where most of the fatalities occurred, a warning system could provide ~20 sec warning (see supplementary material).

Finally, MyShake is first and foremost a seismic network, for which we have developed an early warning algorithm. The network could provide millions of seismic waveforms for a wide range of research activities following large magnitude earthquakes in urban environments. These could be used to generate microseism maps providing information about local amplification effects, and be used to study the impact of the shaking on buildings. The data could also potentially be used to image shallow Earth structure beneath our cities, and perhaps even to image the earthquake rupture process itself.

2.6 Supplementary material

2.6.1 Data collection: The MyShake application

"MyShake" is an android application used to release to private/personal phones. It was released to 75 phones in November 2014 (Figure 2.7). This test release was aimed at student volunteers on the UC Berkeley campus. The trigger algorithm at the time consisted of a simple STA/LTA algorithm [*Allen*, 1978]. The application first requires the phone to remain stationary for 30 minutes, meaning the acceleration is minimal and most likely the phone is siting on a stationary surface. When it meets this requirement, the phone enters into "steady state". The ratio of short-term average (STA) and long-term average (LTA) on any of the 3-components must then exceed a threshold to trigger. When it does, trigger information was immediately sent to CPC including the phone location, time of the trigger, phone ID, and the maximum amplitude. A total of 5 minutes of data was also stored locally on the phone from 1 minute before the trigger to 4 minutes after. A ring buffer stores the last minute of accelerometer data in memory at all times for this purpose. The application also uploads state-of-health (SOH) information every 2 hours and can receive updates and triggers from the CPC. The SOH information provides us with basic information about the number of phones running the application, their location, lifetime of the app, etc. We can also update/change the settings of the application on an individual phone or all phones from the CPC, for example changing the trigger parameters. Finally, we can trigger recording on a phone from the CPC. Either individual phones or the entire network can be triggered to record for a period of time. The waveform data was only uploaded when the phone was plugged into power and had a Wi-Fi connection to minimize power and data-plan usage. All these parameters can be modified remotely. We collected four months of triggered human activity data for our training and testing dataset (Figure 2.7). During this period 17600 triggers (all due to human activities) were uploaded to our CPC.

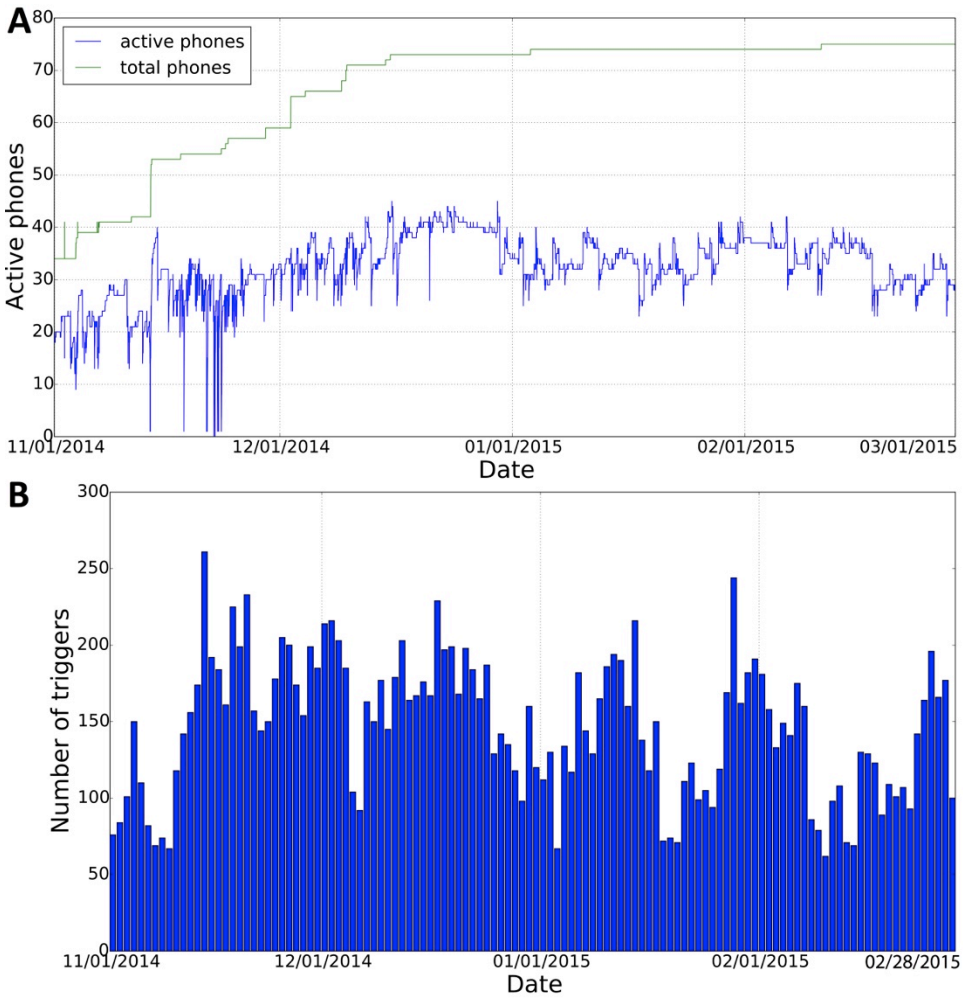


Figure 2.7: MyShake activity November 1, 2014 - February 28, 2015. (a) Number of phones that downloaded MyShake and registered with our network (green curve), and the number of active phones running the application on a given day based on the SOH information (blue curve). Server at CPC restarts during the first month is the reason the number of active phones drops to zero. (b) Number of phone triggers each day with waveforms uploaded to the CPC, a total of 17600 triggers were collected.

Accurate time is key for all data. The drift in the internal clock on the phones is unacceptable for earthquake-related applications, typically ranging from 0.4 – 8.6 sec/day [*Zhong et al.*, 2011]. Thus, geographically distributed nodes need to synchronize their clocks. In the last decade, much research has been conducted to synchronize different internal clocks by referring external signal sources such as power lines, FM radio, Wi-Fi, mobile station, etc. Of them, Network Time Protocol (NTP) is the most commonly used clock synchronization protocol. With a very low network and computation cost, NTP is able to synchronize all the participating nodes within a few milliseconds. In the MyShake application all the accelerometer data is associated with its local device clock, so we synchronize them to Coordinated Universal Time (UTC) via NTP. The MyShake application synchronizes its local clock every 1 hour, thereby

minimizing network and computation cost while ensuring sufficient clock accuracy at all times.

Power usage of the application is also important. Careful selection of which sensors to use and when is needed to reduce the power needs to a level that would not impact normal daily smartphone use. Our goal was an application that could continuously run in the background and still only require the phone to be charged once per day for most/typical phone users. Working within these power requirements, we found that it is possible to monitor the accelerometer data continuously all day. However, it is not possible to continuously use the GPS unit, as it is very power hungry. Instead, we only access GPS at specific times when needed. For the initial 2014 release we only attempted to obtain a location when the phones triggers. When a location request is made to the phone, it returns the best available location. If a GPS location is available it is returned. If not, then the location based on triangulation with cell phone towers is used, if not, then the last available location is used.

The current version of MyShake that we plan to release publically is modified to add the classifier analysis developed to distinguish earthquake from non-earthquake motions, and the use of GPS location has been modified. We continue to have the same initial STA/LTA trigger requirement, after the STA/LTA triggered, we use 2-sec data windows with a 1-sec step to calculate the three key features (IQR, ZC and CAV) up to 10 sec after the STA/LTA trigger. The calculated features in each time window are fed into the Artificial Neural Network (ANN) detector (on the phone) to determine if it is a likely earthquake or not. This two-step approach is implemented so that we do not increase the power requirements, since the STA/LTA method is a simple and low cost computation method. The approach to determining location has also been improved by determining the best available location at the time the phone enters steady state. Now, when the phone enters steady state, the application will try to sample the GPS location. It may take a few seconds to minutes before it gets a stable GPS location. Since phones typically sit in steady state for some time (while sitting on a desk or charging over night) it is unlikely that a trigger occurs in the first few seconds or minutes. If for some reason the phone cannot get the GPS location, e.g. the phone is inside a large building, then the cell phone network location that based on cell phone towers is used. The phone then stores the best available location for the duration of the steady state phase and associated it with the other trigger information when the phone next moves.

2.6.2 Classifier analysis: Detecting earthquakes on a phone

We used three types of data for training, validating and testing our classifier. Firstly, normal human activity data collected from the MyShake November 2014 release for four-month period shown in Figure 2.7. For waveforms to be uploaded, the phone must be stationary, and then move to trigger the STA/LTA algorithm as described above. 10 seconds of data immediately following the human trigger is used in our analysis. We used the first three months of data to train and validate the algorithm, and the last month was kept for final testing.

The second type of data consists of earthquakes recorded on smartphones that were placed on a shake table. These include 241 3-component records from 45 shake table tests runs. The input waveforms into the shake table were past earthquakes with amplitudes rescaled to satisfy the displacement capabilities of the shake table. We only selected the strongest portion of the waveforms recorded by the smartphones, see an example in Figure 2.8. We focus on the strongest portion of the waveforms, as it is difficult for our classifier to distinguish weak earthquake shaking from human activities. This dataset was used entirely for the training and

validation phase.

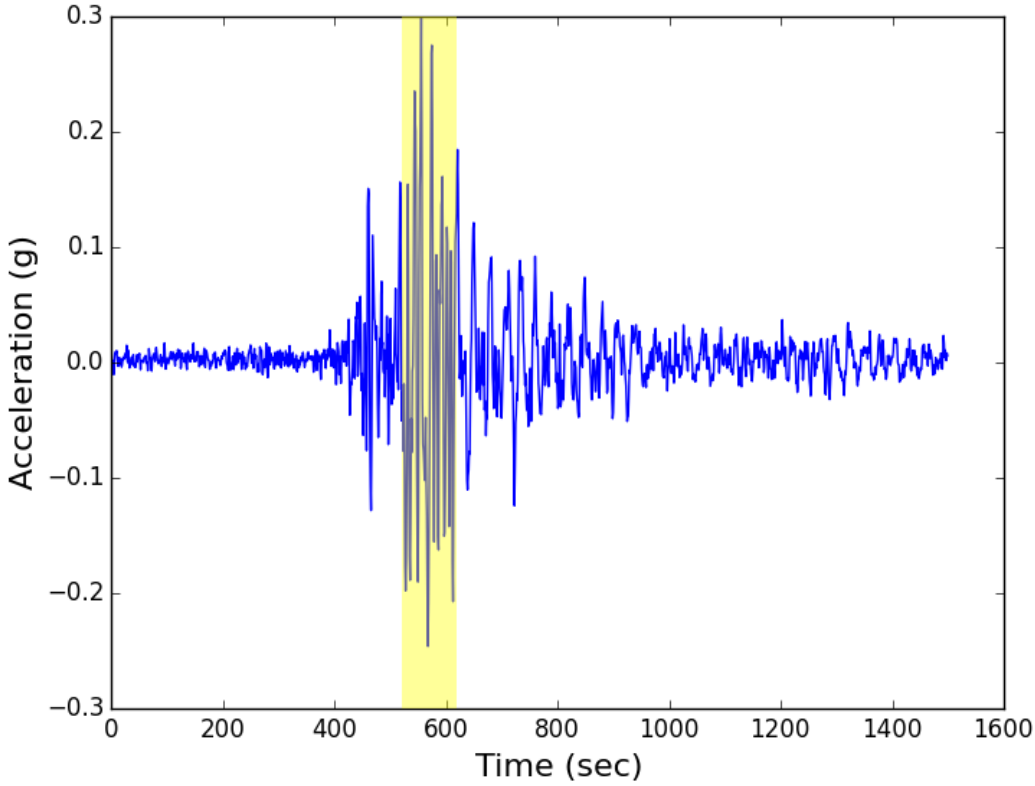


Figure 2.8: Example earthquake record used to train the ANN classifier algorithm. The waveform is the EW component from a regional network station 16.5 km from the epicenter of the western Tottori earthquake (M7.3) of October 6, 2000. The data has been modified to represent a smartphone recording at the same location. Only 2-sec windows of data from the yellow region were used to train our algorithm.

The third type of data also consists of earthquakes, but recorded on regional seismic networks in Japan and US. It was first modified to replicate waveforms recorded on a smartphone. To do this we first converted the 24-bit data to 16-bit data, then we added a smartphone noise record from the noise floor tests to produce accelerometer records similar to what we would record on a phone laying on a sturdy table during the event. Phones are not expected to trigger on the initial low-energy P-waves, especially for smaller earthquakes, instead to trigger on the larger amplitude portions earthquake shaking. We therefore selected windows of data from only the strongest portion of shaking (e.g. Figure 2.8). We used strong motion data from Japan's KiK-Net and K-Net to train and validate our algorithm. Data with horizontal peak amplitude greater than 0.2g for the period from January 1, 1996 to February 1, 2015 was downloaded from NIED (National Research Institute for Earth Science and Disaster Prevention). A total of 317 3-component records from 203 events were selected. To further test the performance of the algorithm, we used earthquake data from the California Integrated Seismic Network (CISN.org). We used 389 3-component records within 30 km of the earthquake

epicenter from 5 events that obtained from CESMD (Center for Engineering Strong Motion Data), NCEDC (Northern California Earthquake Data Center), and SCEDC (Southern California Earthquake Data Center), the results is showing in Table 2.1.

All data was first high-pass filtered in a simulated real-time manner using the method described in [*Kanamori et al.*, 1999]. A range of characteristics in overlapping 2-sec data windows was calculated with 1-sec step. We experimented with using different window lengths and steps and found this to be the best compromise between having more data and keeping the window short to detect earthquakes more rapidly. 18 different features including frequency features, amplitude features, and statistical features was tested, the approach similar to the ones described in [*Kong and Zhao*, 2012]. All features had low to moderate computational requirements making it feasible to rapidly determine their values on a phone.

Since there were far more data points from human activities than that from earthquake data, this imbalance of classes could affect our classifier. In order to create a dataset with equal classes, we used the kmeans cluster method [*Kuhn and Johnson*, 2013] to group the human activities into a number of clusters, with the number of clusters being equal to the number of the earthquake data points. The centroid of the cluster was taken to represent human activity data. This not only created a balanced dataset for us to train our classifier, but also reduced the computation burden during the training.

We selected the best 3 features to distinguish between earthquake and non-earthquake data using greedy forward feature selection. They are the interquartile range (IQR) between the 25th and 75th percentile of the acceleration vector sum, the zero crossing rate from the component with the highest value (ZC), and the cumulative absolute velocity (CAV) of acceleration vector sum. IQR is an amplitude parameter that shows the middle 50% range of amplitude of the movement. ZC is a simple frequency measure. CAV is a cumulative measure of amplitude on the three components in the time window and is determined as in equation 2.1.

An ANN (artificial neural network) approach is used to classify a particular data window as an earthquake or not an earthquake. Each feature was first scaled to a range of 0 to 1. We used an ANN with one hidden layer and completed a grid search to test different numbers of neurons. Best performance is achieved when the ANN has 1 hidden layer with 5 neurons (Figure 2.9) with a standard sigmoid activation function defined as:

$$s(x) = \frac{1}{1 + e^{-x}}$$

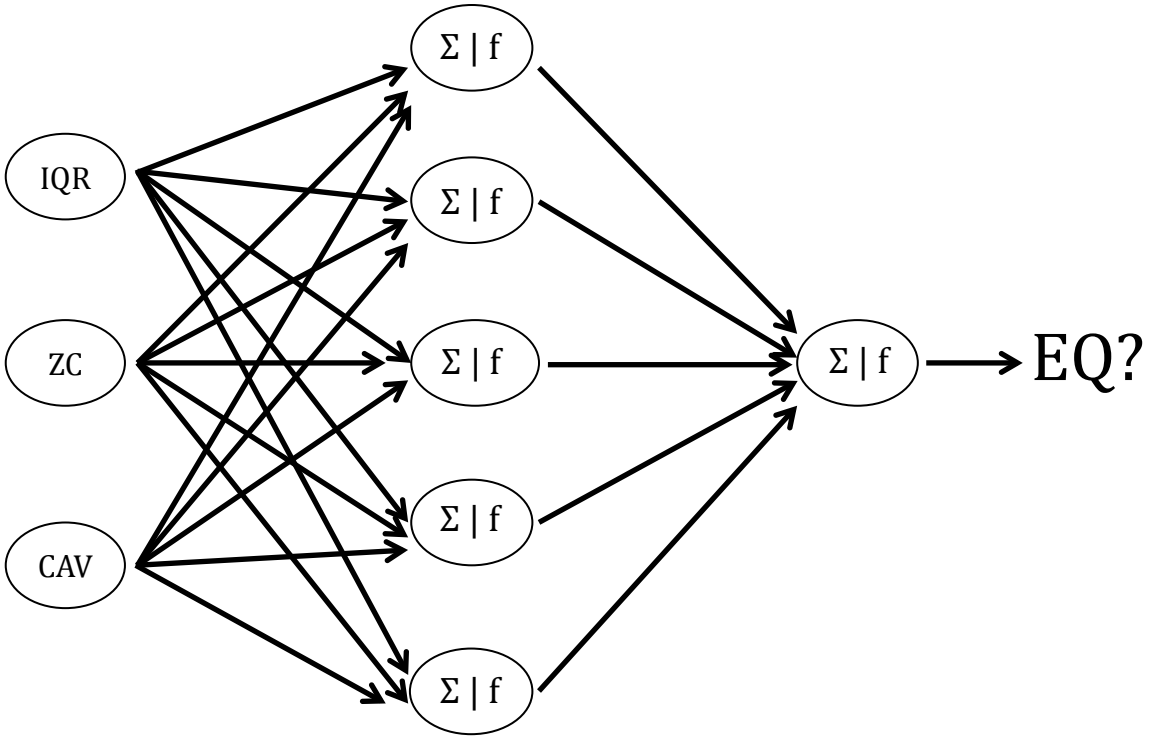


Figure 2.9: Structure of Artificial Neural Network (ANN) classifier algorithm. It has three layers: one input layer with 3 nodes, a hidden layer with 5 nodes, and an output layer with 1 node. For the hidden layer and output layer, the inputs from the previous layer to the each node will be first summed and then fed into an activation function shown as f.

The ANN was trained and validated using 3 months of human activity data, and earthquake data from shake table tests and Japanese events. The dataset was split multiple times using 70% of the data for training and 30% for testing for cross-validation tests. The accuracy of the classifier when applied to the test datasets is very good, showing 98% to 99% accuracy each time (Table 2.2, and Figure 2.10).

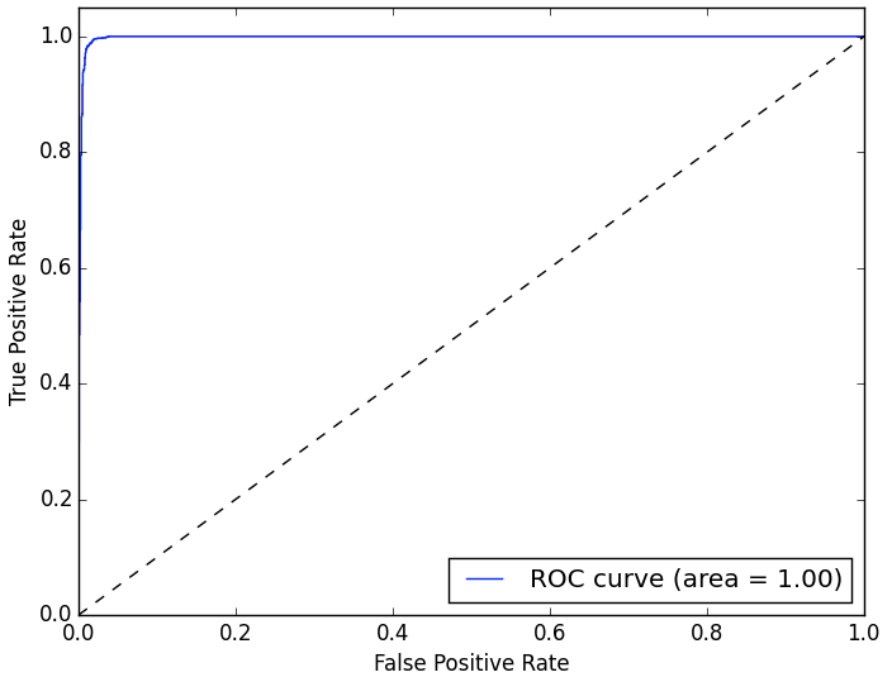


Figure 2.10: Receiver operating characteristic (ROC) curve. Shows the ANN classifier performance on 30% test data split from the training data. The ROC curve shows the false positive rate (classified as earthquake when it is a non-earthquake) on the x-axis, , against true positive rate (classified as an earthquake when it is an earthquake) on the y-axis. Ideally, the curve will climb quickly toward the top-left corner meaning the model correctly predicted the cases. Our result is quite close to the ideal cases.

	1	2	3	4	5	6	7	8	9	10
Score	0.9893	0.9830	0.9839	0.9811	0.9919	0.9919	0.9893	0.9857	0.9821	0.9966
Mean	0.986 (±0.001)									

Table 2.2: Accuracy score for ANN classifier with 10-fold cross validation. The score row shows the accuracy score for each run defined as

$$accuracy(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1} I(\hat{y}_i = y_i),$$

where n is the number of samples used, and *I* is a function that takes 1 when the argument is true, and 0 otherwise. This means that if the ANN classifier correctly classify the data, then *I* will be 1, otherwise 0. So the higher the average score, the better the ANN classifier. We ran a 10-fold cross validation, which means we split data into 10 sets of n/10 and trained on 9 datasets and tested on 1 dataset. We repeat this process 10 times, and each time select a different dataset as

test set. The mean row shows the average score from the 10 runs, with the deviation showing in the parentheses.

Our trained ANN classifier algorithm was tested by applying it to a dataset consisting of data that was not used in the training/validation process. This contained the last month of MyShake human activity data (February 1 to 28, 2015), and data from large US earthquakes modified to represent waveforms recorded on smartphones. Note that no selection criteria were applied to the US earthquake data (recall that for the Japan earthquake data, we only selected the stations have clear large amplitudes). We applied the classifier to all available waveforms, and the results of this validation are shown in Table 2.1 and described in the main text.

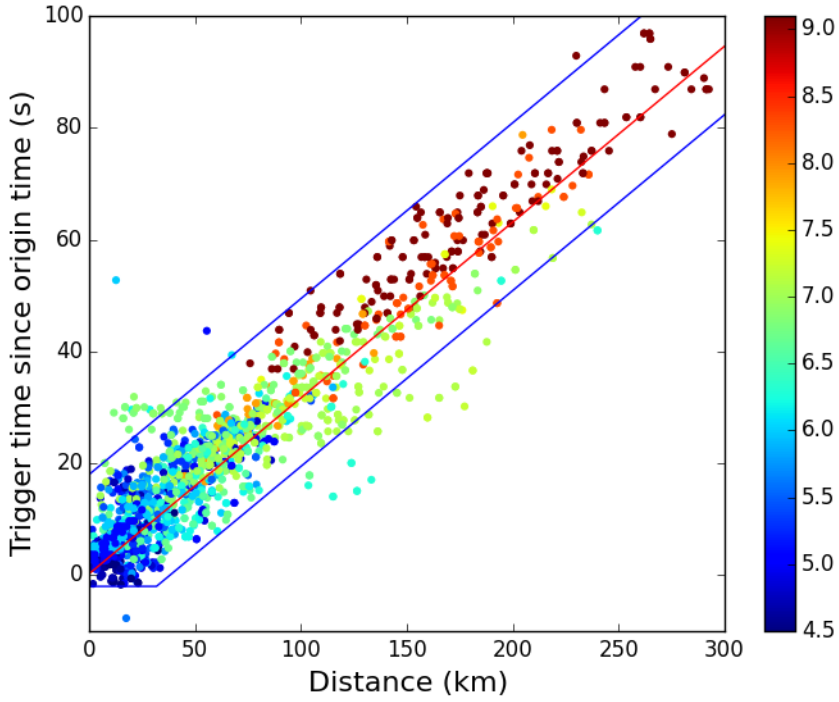


Figure 2.11: Phone trigger times versus epicentral distance. The regional network data from California and Japan was modified to phone-quality data and then our classifier applied to the data to determine when a trigger occurs. The red line is the best-fit to the data and has a moveout velocity of 3.2 km/sec; most triggers are generated by the S-wave or the later surface wave. The blue outline is the time-space window used for association of triggers with an event by the network detection algorithm.

2.6.3 Network detection algorithm

Our first-generation network detector identifies multiple triggers in a space-time cluster, and is based on the approach used in our ElarmS-2 earthquake early warning algorithm for traditional regional seismic networks [*Kuyuk et al.*, 2014]. We stored triggers for 20 seconds and look for 4 or more triggers within a 10 km radius region that can be associated. We require greater than 60% of operating phones to have been triggered within 10 km of the location of the event for an event to be declared (the estimated event location is the centroid of the locations of

the triggered phones.). The origin time is assumed to be that of the first phone to trigger. The magnitude is estimated based on the peak ground acceleration of the triggered phones as described in the main text. Triggers from phones at greater than 10 km must then fall within a defined space-time region to be associated with the event (Figure 2.11).

Earthquake	Origin time	Event latitude	Event longitude	Alert time	Mag.	Location error (km)	Origin time error (sec)
La Habra: True	March 29, 2014 04:09:42	33.932	-117.917		5.1	3.76	2
La Habra: Estimated	04:09:44	33.900	-117.930	04:09:47	5.2		
Parkfield: True	Sep 28, 2004 17:15:24	35.815	-120.374		6.0	1.55	2
Parkfield: Estimated	17:15:26	35.810	-120.390	17:15:28	5.5		

Table 2.3: Simulated network detection performance for US earthquakes. Simulated phone triggers from 2014 La Habra M5 earthquake and 2004 M6 Parkfield earthquake were used to test the network detection algorithm. The magnitude, location and origin time estimates and errors are given for the initial MyShake estimates.

We used simulated phone triggers from two earthquakes to test the performance of the algorithm: 2014 La Habra M5 earthquake and 2004 M6 Parkfield earthquake. Table 2.3 for the performance of the algorithm. In these simulations we assume zero latency due to processing and network transmission. We estimated the actual latency that will be introduced into the system due to the processing on the phone and network transmission. First, to estimate the processing delay of the ANN on the phone we did a test run for one night and found the average processing time is to be 4.5 milliseconds. Second, the transmission of the trigger data from phone to CPC is via UDP (User Datagram Protocol), which is a common choice for time-sensitive applications. We found that the average delay time of transmitting the data from the phone to the CPC via UDP is 50 milliseconds.

In addition to the simulated phone triggers from real earthquakes, we generated phone-triggers for a simulated network to test performance sensitivities of our network detector. We used a 1° by 1° box and randomly distributed N stations within the box where N can be 100, 200, 300, 400 or 500. We allowed randomly distributed false triggers at a rate based on the assumption that 10% of phones initially trigger due to movement every second, and then 7% are classified erroneously as an earthquake. We then added earthquake triggers due to earthquakes using the following method.

The trigger time for each phone is based on Figure 2.11. Given the distance of the phone from the epicenter, the trigger time is randomly selected within the time range given by the blue

lines on Figure 2.11. To determine a probability that a phone triggers, we developed a simple regression relation for the probability of a trigger given the estimated peak ground acceleration (PGA) at the site. We estimated peak ground acceleration at the site using a standard ground motion prediction equation [*Snoke*, 2009]. Our observations from the M5.1 La Habra earthquake are that the probability a phone triggers is 1, 0.8, 0.4, 0.25, 0.1, 0.01 at epicentral distances up to 5, 10, 20, 30, 40, and 50 km respectively. Using these observations we performed a simple regression between $\log_{10}$PGA and trigger probability. The resulting regression relation is

$$P = 0.798 \times log_{10} PGA - 0.557$$

where P is the probability that a phone is triggered. In the case that P>1 we set P=1 and P<0 we set P=0.

Number of stations	Location error (km)	Origin time error (sec)	Detection time after true origin time (sec)	Events not detected (out of 1000)
N = 100	14.02±8.92	4.41±2.80	6.59±2.87	11
N = 200	5.29±4.42	1.77±0.96	3.93±0.99	32
N = 300	4.36±4.79	1.42±0.77	3.53±0.80	0
N = 400	3.56±3.18	1.27±0.66	3.48±0.69	0
N = 500	3.50±3.86	1.26±0.73	3.51±0.63	0

Table 2.4: Simulated network performance for various phone densities. N is the number of randomly distributed stations within a 1°x1° box (~111x111 km); we did 1000 simulations in each case for a M6.0 earthquake. The location errors are the differences between the true earthquake location and the estimated earthquake location. The origin time errors are the time difference between the true earthquake origin time and that estimated. The detection time is the time after the true earthquake time that the algorithm detects it. In all cases we show the average value ± standard deviation. The last column shows the number of simulations in which the earthquake was not detected.

In 1000 simulations for each value of N, there were no false network earthquake detections. For N=500, 400 or 300 the performance is similar with all events detected ~3.5 sec after the origin time with location errors of ~4km (Table 2.4). For N=200, 32 of the 1000 events were not detected, and 11 were not detected for N=100 (Table 2.4). It also took longer to detect the events, and the locations had larger errors for N=100 and 200 illustrating the need for a dense distribution of smartphone detectors for this approach to work. The N=300 case corresponds to average distance between phones of 6.4 km. We also did 1000 simulations with only noise data without earthquakes, and found the algorithm did not have false alert issued. This is due to the requirement that >60% of active phones trigger within a 10 km radius for an earthquake to be declared.

2.6.4 Estimate warning time for Katmandu, Nepal

For the M7.8, 25 April 2015 earthquake in Nepal we can estimate the possible warning time in Katmandu using our smartphone seismic network approach. The location of the epicenter is 28.147°N, 84.708°E, and the location of Katmandu is 27.700°N, 85.333°E, a separation of 79 km. The S phase of the earthquake will arrive at Katmandu in 25.2 seconds based on iasp91 model (*39*). Assuming there are smartphones near the location of the earthquake, and because our network detection algorithm makes use of phones within 10 km of the epicenter, we would expect an earthquake detected when the S-wave reaches 10 km from the epicenter, which is 3.9 seconds after the origin time based on iasp91. Therefore, there could be ~20 seconds warning if we have a smartphone seismic network in Nepal.

2.7 Acknowledgments

We thank Deutsche Telekom's Silicon Valley Innovation Center and SenseOS for technical support. We also thank The Pacific Earthquake Engineering Research Center (PEER.org) provided us with access to their shake tables. Data was obtained from the Northern and Southern California Earthquake Data Centers (NCEDC.org, SCEDC.org), CESMD, and the NIED for Japanese events. This work is funded by Deutsche Telekom's Silicon Valley Innovation Center and UC Berkeley.

Chapter 3 Initial Observations from MyShake Network

Published as: Qingkai Kong, Richard M Allen, Louis Schreier. 2016, "MyShake: Initial Observations from a Global Smartphone Seismic Network", Geophysical Research Letters, doi:10.1002/2016GL070955

3.1 Abstract

In the first 6 months since the release of the MyShake app, there were almost 200,000 downloads. On a typical day about 8,000 phones provide acceleration waveform data to the MyShake archive. The on-phone app can detect and trigger on P-waves and is capable of recording magnitude 2.5 and larger events. More than 200 seismic events have been recorded so far, including events in Chile, Argentina, Mexico, Morocco, Nepal, New Zealand, Taiwan, Japan, and across North America. The largest number of waveforms from a single earthquake to date comes from the M5.2 Borrego Springs earthquake in southern California, for which MyShake collected 103 useful 3-component waveforms. The network continues to grow with new downloads from the Google Play store every day, and expands rapidly when public interest in earthquakes peaks such as during an earthquake sequence.

3.2 Introduction

Since the introduction of low-cost accelerometers in consumer devices such as cars and computers, seismologists have been experimenting with how these sensors might contribute to the science of seismology and hazard reduction [*Allen*, 2007; *Cochran et al.*, 2009b; *Fleming et al.*, 2009; *Chung et al.*, 2011; *Clayton et al.*, 2011, 2015; *Wu et al.*, 2016, 2013; *Evans et al.*, 2014; *Wu and Lin*, 2014; *Wu*, 2015]. While these devices have significantly lower price tags than traditional seismic stations, the data is lower quality and the operation of networks of low-cost devices is complex and not necessarily low-cost. Various types of lower-cost sensor networks have been explored with varying degrees of success, e.g. *Allen* [2012]. At the higher-quality end of the spectrum, the USGS Netquakes devices include a high-quality MEMS accelerometer in a station package that is installed by engineers in household basements and makes use of the in-home wifi [*Luetgert et al.*, 2009, 2010]. Other efforts have made use of USB accelerometers attached to personal computers or low-cost computers as with the Community Seismic Network [*Clayton et al.*, 2011, 2015; *Kohler et al.*, 2013] and Quake Catcher Network [*Cochran et al.*, 2009a, 2009b; *Chung et al.*, 2011; *Lawrence et al.*, 2014]. In all these cases hardware must be transported from the network operator to a station host. Both hardware and software must then be installed and maintained for the station and network to continue to function.

The advantage of using smartphones is that all the hardware is already packaged in a device that is ubiquitous in urban environments around the world. In addition, convenient software distribution and maintenance platforms exist in the form of the Google Play and iTunes stores and the associated software development kits. The disadvantages of smartphones as seismic sensors are also obvious: the phones are not fixed, phone resources are not tailored for seismology, recording earthquakes is not typically a priority for owners, the phones experience all kinds of motions that have nothing to do with earthquakes, and rapid full waveform data

recovery may present challenges.

Multiple efforts are underway using smartphones to detect earthquakes. *Bossu et al.*, [2015, 2016] describe the collection of quick eye-witness reports from people within a few tens of minutes after a felt-earthquake occurrence. Multiple efforts make use of the accelerometer. In some cases, a "trigger" message is generated and sent to a central server when a phone moves [*Faulkner et al.*, 2011; *Olson et al.*, 2011; *Finazzi*, 2016], in others phones are used in a dedicated way to record earthquake shaking by attaching them to walls or other structures and recording continuously [*Naito et al.*, 2013]. The GPS/GNSS sensor on the phone can also be used to detect ground motion when the motion is sufficiently large [*Minson et al.*, 2015]. MyShake attempts to combine all of these elements by turning a typical personal smartphone into a seismometer. The MyShake app has a filter to distinguish earthquakes from other human activities, it uploads earthquake triggers to a real-time server for analysis, and also uploads the acceleration timeseries data to a server for further research analysis [*Kong et al.*, 2016a].

This chapter presents initial observations from the seismic data recorded by MyShake since the public release in February 2016. We detail the rapid expansion to a global seismic network recording earthquakes across six continents, and show that smartphone sensors are capable of recording seismic events with magnitude 2.5 and larger. For larger earthquakes, these sensors can record the entire wave train starting from P wave. Previous shake-table tests have assessed the quality of the smartphone waveform data [*Dashti et al.*, 2011, 2012; *Reilly et al.*, 2013; *D'Alessandro and D'Anna*, 2013; *Kong et al.*, 2016], here we compare smartphone recordings in the field to nearby traditional seismic stations. Finally, we examine the potential for MyShake to provide seismic data in many regions where there are little or no traditional seismic stations such as Nepal and Ecuador, and where seismicity is a relatively new phenomenon like Oklahoma.

3.3 MyShake Methodology Refresh

MyShake was developed on the android platform as an application to monitor the accelerometers inside the smartphones. The motion of the smartphones is summarized into 3 key parameters that feed into an Artificial Neural Network (ANN), which has been trained to distinguish earthquakes from human activities. Once the ANN algorithm detects an earthquake-like motion, the app will send a message in real-time to the server. This message, which contains location, time, and amplitude of the trigger, can be used for earthquake early warning or other types of rapid detection applications. At the same time, the app collects 3-component acceleration time-series data, at 25 Hz. The waveform data has a duration of 5 minutes, including 1-minute before, and 4-minutes after the trigger. The one minute of data before the trigger ensures that the entire earthquake waveform is recorded even when the phone only triggers on a later phase of the ground motion. When the phone is connected to WIFI and power, the waveform recordings are uploaded to the server for further analysis. For more details about the app and methodology, see *Kong et al.*, [2016].

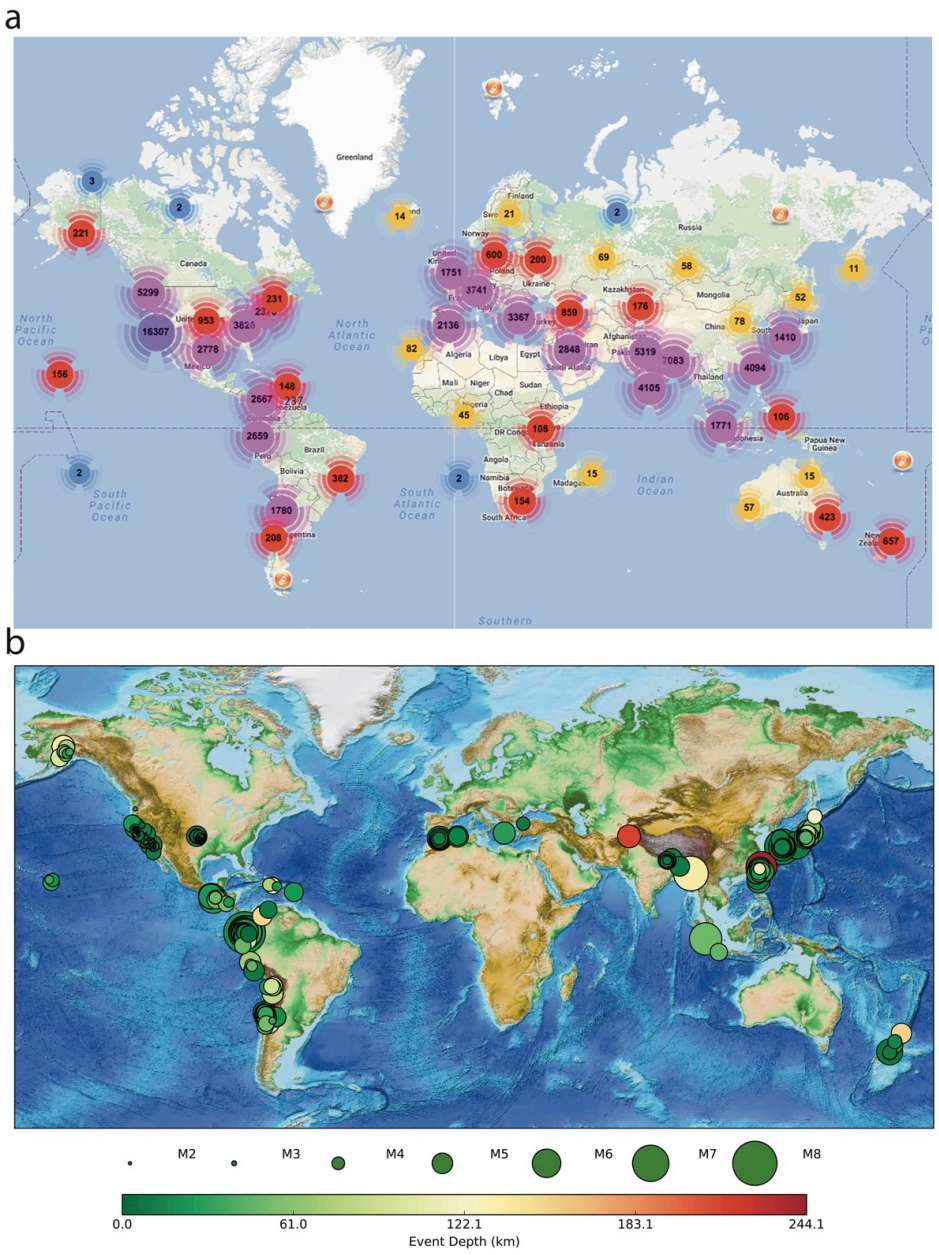


Figure 3.1: Distribution of MyShake registered users and detected earthquakes. (a) Registered MyShake users are shown in clusters. The number in each circle indicates the number of registered users in the cluster, and the color of the circle shows the order of the number of phones, i.e., a purple circle indicates the number of phones is of order tens of thousands, magenta is thousands, red is hundreds, yellow is tens, and blue for less than 10. (b) 237 Earthquakes recorded by MyShake users since February 12, 2016. The locations of the earthquakes are shown as circles, which are color-coded by the depth and whose sizes are scaled by the magnitude of the earthquake. Figures are generated on August 11th 2016.

Since MyShake was released publicly on February 12, 2016 in the Google Play store, there have been almost 200,000 downloads and the app is presently installed on 36,000 phones

distributed across six continents (numbers from Google Play Store) within 6 months. Figure 3.1a shows the global distribution of all phones registered with the system. The number of phones contributing shows peaks in North America and some other places where earthquake hazard is high including Nepal and India. On a typical day between 8,000 and 10,000 phones provide data to the system (see Figure 3.5).

3.4 Seismic data recorded by MyShake

As of August 11th, 2016, MyShake has recorded 237 earthquakes from M2.5 to M7.8 around the globe since released to the public ("recorded earthquakes" are defined as having at least one good smartphone-recorded earthquake waveform sent back to the server, and confirmed by a seismologist). Figure 3.1b shows the distribution of the recorded earthquakes, and includes regions with good seismic network coverage like the U.S., Taiwan, Japan, and Chile, and also areas without dense networks, such as Nepal and Ecuador. Not surprisingly, locations with higher density of MyShake users, like California, have a larger number of earthquake detections. This is due to the higher density of users closer to the epicenter. Example waveforms from around the world are included in Figure 3.6. Not only shallow, but also large deep earthquakes have been detected by MyShake. These are shown by the warmer color circles in the Figure 3.1b. The distribution of the magnitude and depth of the recorded earthquakes can be found in Figure 3.7.

The earthquake which has generated the most waveforms to date is the M5.2 earthquake that occurred near Borrego Springs in southern California on June 10, 2016 at 08:04:38 UTC. Figure 3.2a shows the status and performance of the MyShake network at the time of the event. The green dots show the locations of the phones that triggered on the earthquake ground motion using the ANN algorithm on each phone. The orange dots are the phones that were "ready", i.e. they were monitoring the accelerometer to detect an earthquake, but did not trigger on this event. The red dots show other MyShake phones that were in communication with the network but were not monitoring for an earthquake at the time of the event. Not surprisingly, the percentage of phones that triggered on the event decreases with increasing epicentral distance as the amplitude of the ground motion decreases.

The trigger times of the phones is shown in Figure 3.2b along with the expected P- and S-wave arrival times. Most phones are triggered on the P- or S-waves, as would be expected, and it is encouraging to see that the ANN detection algorithm is still recognizing the earthquake out to distances of ~200 km.

Figure 3.3a is the record section showing horizontal component waveform data from phones out to 200 km from the epicenter. The S-wave energy is clearly recorded by the smartphone sensors out to 200 km, and the P-wave energy is also visible on some phones at these distances. The P-wave signal is clearer on the vertical component record section (**Figure 3.8**).

One of the key parameters for earthquake hazard studies is the Peak Ground Acceleration (PGA). Figure 3.3b presents a comparison of the PGA values observed by MyShake phones and traditional seismic stations. While the values are similar and show the same trend with epicentral distance, the ratio of the PGA values from the phones to that of the nearest seismic station is 2.0. Figure 3.9a shows a histogram of PGA difference between MyShake recordings and the closest traditional seismic station. This may reflect the fact that the phones are in buildings and on tables rather than being free field sites as with most traditional stations. It may also reflect the

fact that most people (and their phones) live in basin locations and on sediments leading to amplification effects. Figure 3.9b shows a comparison of the occurrence time of the PGA value on MyShake recordings and traditional seismic stations showing correlation with the S-wave arrivals. Figure 3.4a and b show comparisons of the waveforms recorded on a smartphone and a nearby traditional seismic station. It shows good agreement between the waveforms (they are separated by 1-2km) but also shows that the PGA is greater on the phone records. For other components comparisons see Figure 3.10 and Figure 3.11.

The other recorded earthquakes typically have far fewer waveforms than the Borrego Springs earthquakes, because of lower density of MyShake phones. The largest earthquake recorded to date is the M7.8 2016-04-16 Ecuador earthquake. In this event two phones triggered at distances of 170 and 200 km (The location of the phones and earthquake is shown in Figure 3.12a). Figure 3.4c is the waveform record at 170 km and shows that the phone triggered on the P-wave arrival even at this great distance. Shortly following the S-wave arrival there is a very large acceleration likely due to the phone owner picking up the phone (see the whole waveform in Figure 3.12).

MyShake has also recorded multiple earthquakes in Oklahoma. Figure 3.4d is one example from the 2016-02-13 M5.1 event recorded at a phone 130.5 km away. While the individual counts are clearly visible in the record, the phone still triggered on the P-wave arrival. Figure 3.13 shows the map of this earthquake, and the waveforms for the 3 components. Additional examples of P-wave recordings are shown in Figure 3.14 illustrating that P-wave arrivals are typically recorded out to ~100 km for M5 and larger events. Figure 3.6 shows some more examples of recordings from other regions.

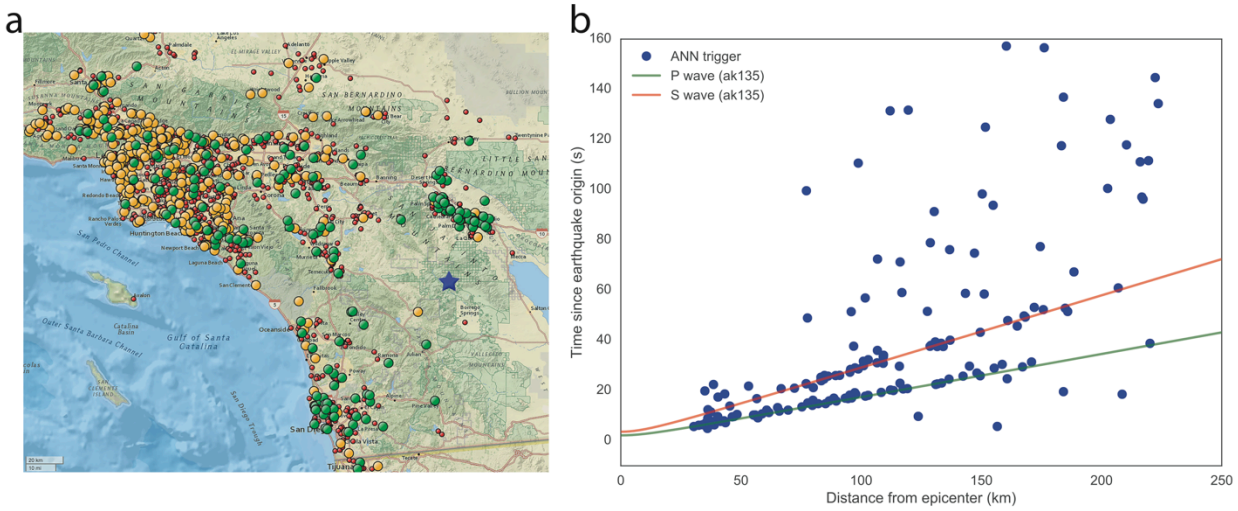


Figure 3.2: (a) Location of the M5.2 Borrego Springs earthquake and the MyShake phones at the time of the event. Blue star is the epicenter of the earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were ready to detect the earthquakes but did not. (b) MyShake trigger time vs distance. Blue dots are the phones' trigger times, and the green and red curves are the estimated P and S wave travel time based on Model ak135 [*Kennett et al.*, 1995].

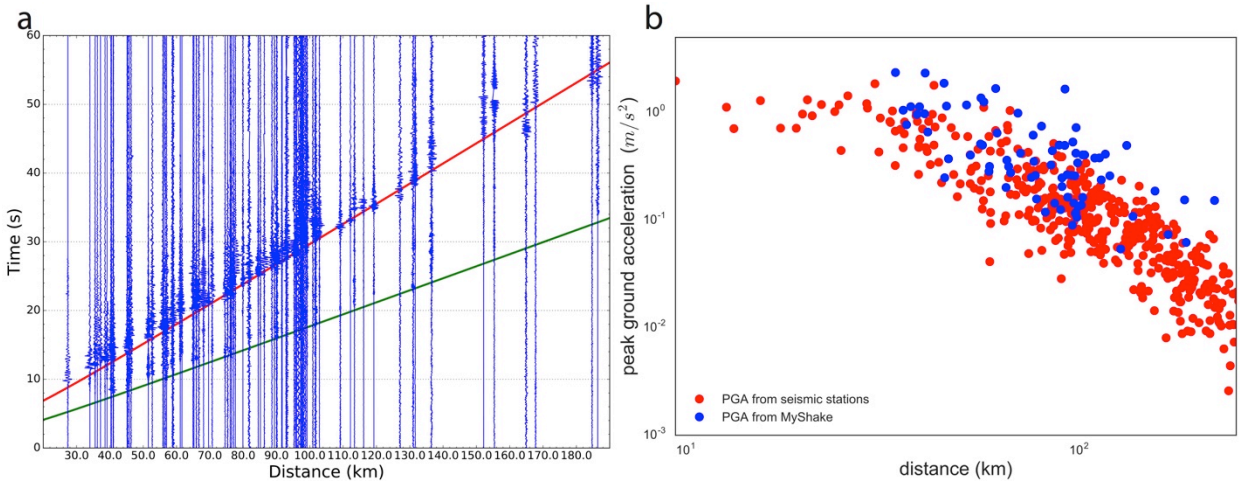


Figure 3.3: (a) Record section plot for phones within 200 km. Each blue trace is one horizontal recording from MyShake user, and the green and red curve is the estimated P and S wave based on ak135. Amplitudes of the recordings are normalized in each trace. (b) PGA value observations with distance. PGA values from MyShake (blue) and traditional seismic stations (red) are shows as observed on the largest horizontal component. The seismic station data are from Southern California Earthquake Data Center (SCEDC).

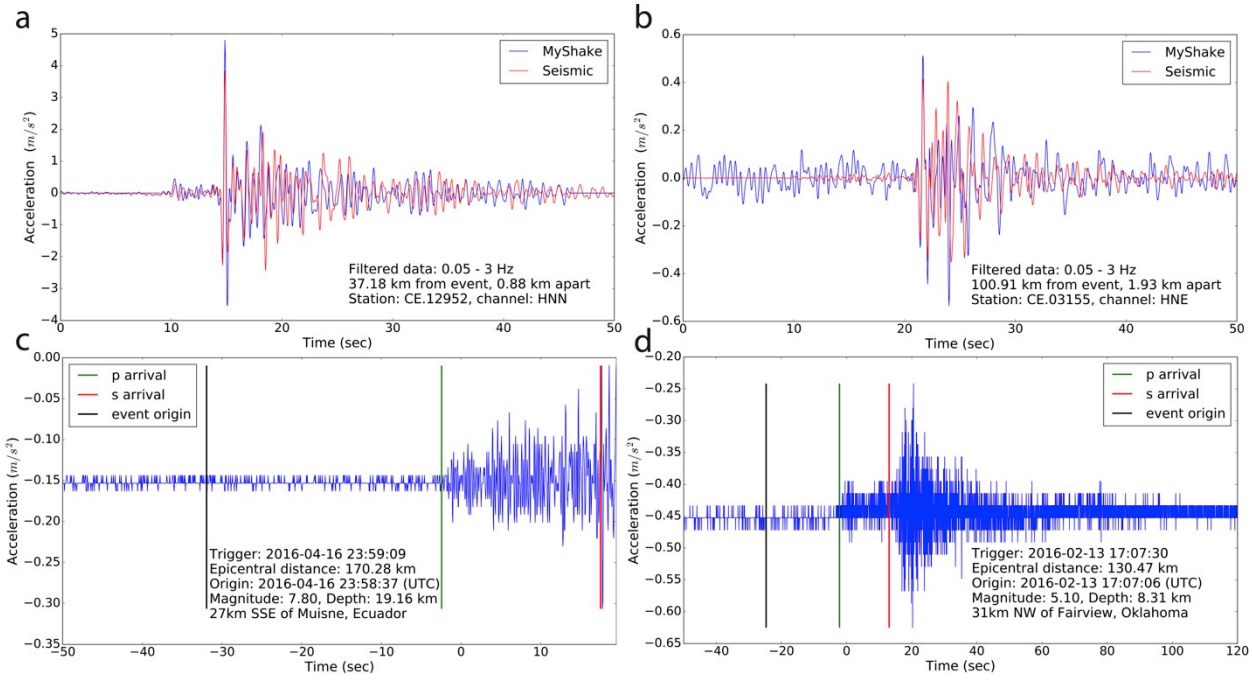


Figure 3.4: Example MyShake waveforms. (a and b) Comparison of the waveforms recorded by MyShake and a nearby traditional seismic station (horizontal component) for the M5.2 Borrego Springs earthquake. (a) MyShake waveform recorded 37.2 km from the epicenter, and a traditional seismic station 0.88 km from the smartphone. (b) MyShake waveform recorded at 100.9 km from the epicenter, and a traditional seismic station 1.93 km from the smartphone. See **Figure 3.10** and **Figure 3.11** for the comparison of other components. (c) M7.8 Ecuador

earthquake recorded at 170 km away from epicenter. (d) M5.1 Oklahoma earthquake recorded by a phone at 130 km from epicenter. For (c and d) the zero time is the phone trigger time. The vertical black, green and red lines are the origin time and predicted P- and S-wave arrival time respectively (estimated using ak135 model).

3.5 Discussion and conclusions

Based on the data recorded by the MyShake network during the first 6 months of operation, the MyShake approach of using personal smartphones has the clear potential to provide data useful for multiple seismological studies and hazard reduction efforts. It is remarkable how quickly the network has grown, primarily driven by news media and citizen scientist interest in the project. While technical challenges still remain, and there are opportunities to improve the data quality through further improvements in the on-phone app, the network is already collecting a substantial volume of earthquake timeseries. MyShake has already provided data from many earthquake-prone areas, including the US, Nepal, Chile, Japan, Taiwan, New Zealand.

The MyShake data is of sufficient quality to be useful in many types of scientific and hazard reduction projects. The data collected shows that the full seismic waveform (P-wave, S-wave and surface waves) can be recorded with a high signal to noise ratio at distances in excess of 100 km for earthquakes of magnitude 5 and larger (see examples in Figure 3.15 and Figure 3.16). It can also provide peak ground motion information for much smaller earthquakes; a M2.5 earthquake is the smallest detected to date. The network can provide a very dense array of stations across urban centers if deployed on enough phones, providing a perhaps unprecedented opportunity for full wavefield analysis. The network can also be used for more traditional seismic studies as it can be used to detect, locate, and estimate the magnitude of earthquakes in regions that have few or no seismic stations. Micro-zonation peak ground motion maps can also easily be generated from the data. These maps may also have a third dimension when arrays of phones also provide observations on multiple floors of high-rise buildings. More events with large numbers of observations are needed to fully understand the potential uses of this data.

The rapid expansion of the network and the large number of recordings are all possible because we harness a ubiquitous hardware/software package: android smartphones. While the sensor network therefore already exists—it is estimated that there are over one billion worldwide—the challenge is in reaching enough sensor owners and persuading them to run MyShake. MyShake must therefore minimize any interference with other phone functions. This means that the applications must run in the background and consume as little power as possible. We must also provide the owner with some value. The current version of the app has a user interface that provides basic information about “recent” (past week) earthquakes as many other apps do. In addition, there is some educational material with information about past earthquakes including videos illustrating the intensity of shaking at the users’ location in those past earthquakes, and information on how to be earthquake safe. Finally, users are participating in a citizen science project whereby they are helping to develop and test the MyShake network. Over the coming months and years substantial effort will be needed to maintain and grow the number of users.

One key objective is also to use the network to deliver earthquake alerts as described by

[*Kong et al.*, 2016a]. This is not only the right thing to do as it will reduce the impact of earthquakes on MyShake participants, but it is also important for MyShake to provide this service in order to increase that value of the app to phone owners and thereby increase the number of phones running MyShake and recording earthquakes. The fact that we can clearly detect P-wave arrivals with MyShake will allow the application of P-wave based methodologies for early warning as are currently employed by traditional seismic networks running early warning systems [*Allen et al.*, 2009b].

Perhaps the most important conclusion is that MyShake has demonstrated the potential to collect seismic waveform data of similar density and useful quality to what we are accustomed to in California, Japan and the few other densely instrumented regions. The network can provide more data in areas with few seismic stations like Nepal, Ecuador, Haiti, etc., including regions with few stations because significant-risk seismicity is a new phenomenon like Oklahoma and Texas.

3.6 Supplementary material

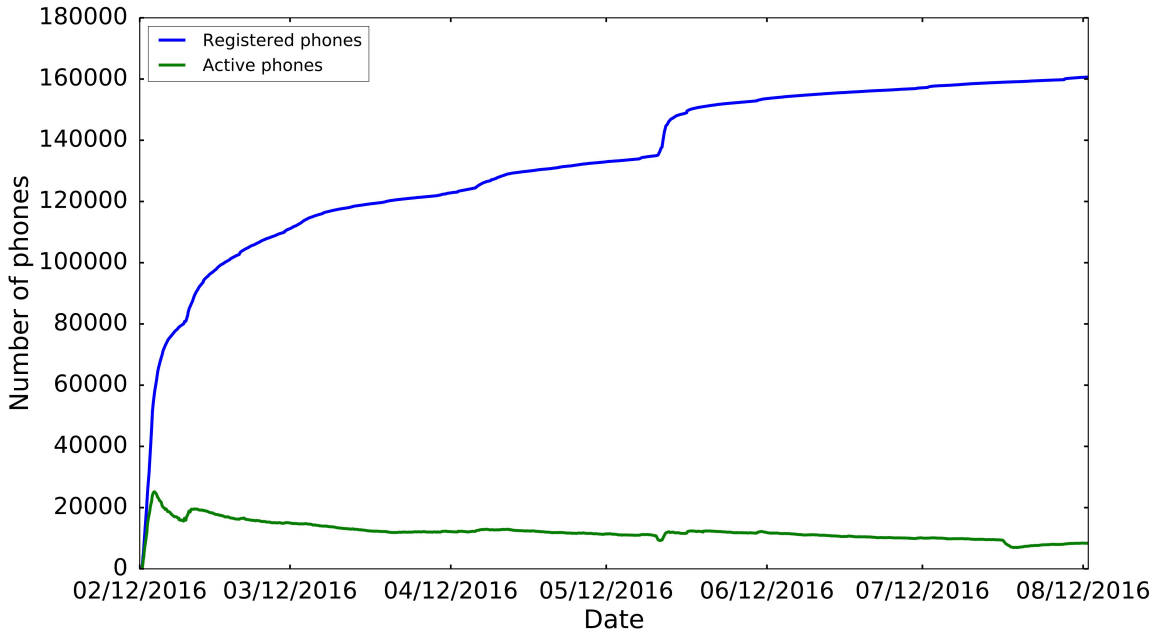


Figure 3.5: The time history of the phones registered with our server (blue curve), and the phones actively contributing data to MyShake in a 24-hour interval (green curve). Registered phones are defined as the phones with MyShake installed that have sent at least one data point to our server. The total number of registered phones is a little lower that the total number of phones that have downloaded and installed the app. Active phones are defined as the phones sent data back to the server within last 24 hours.

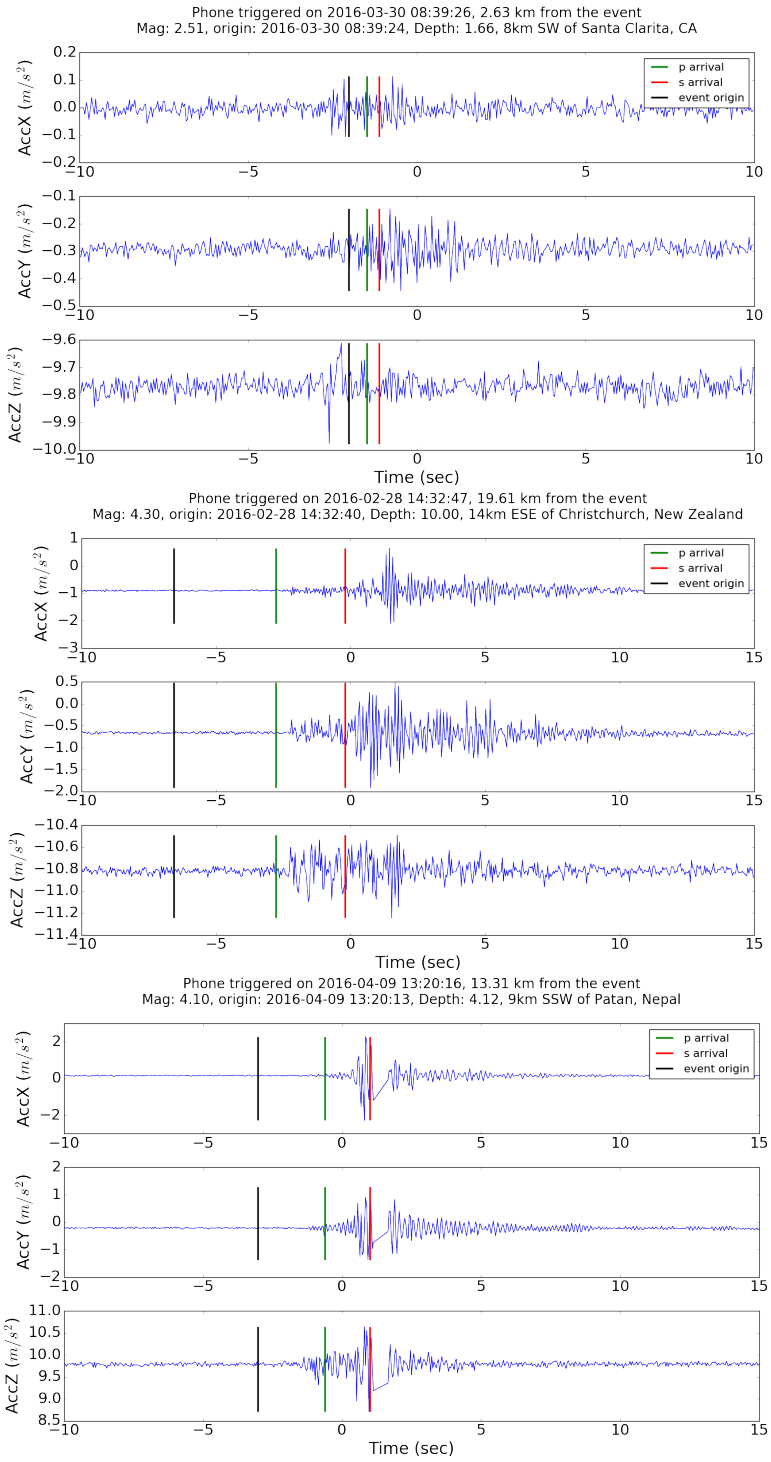
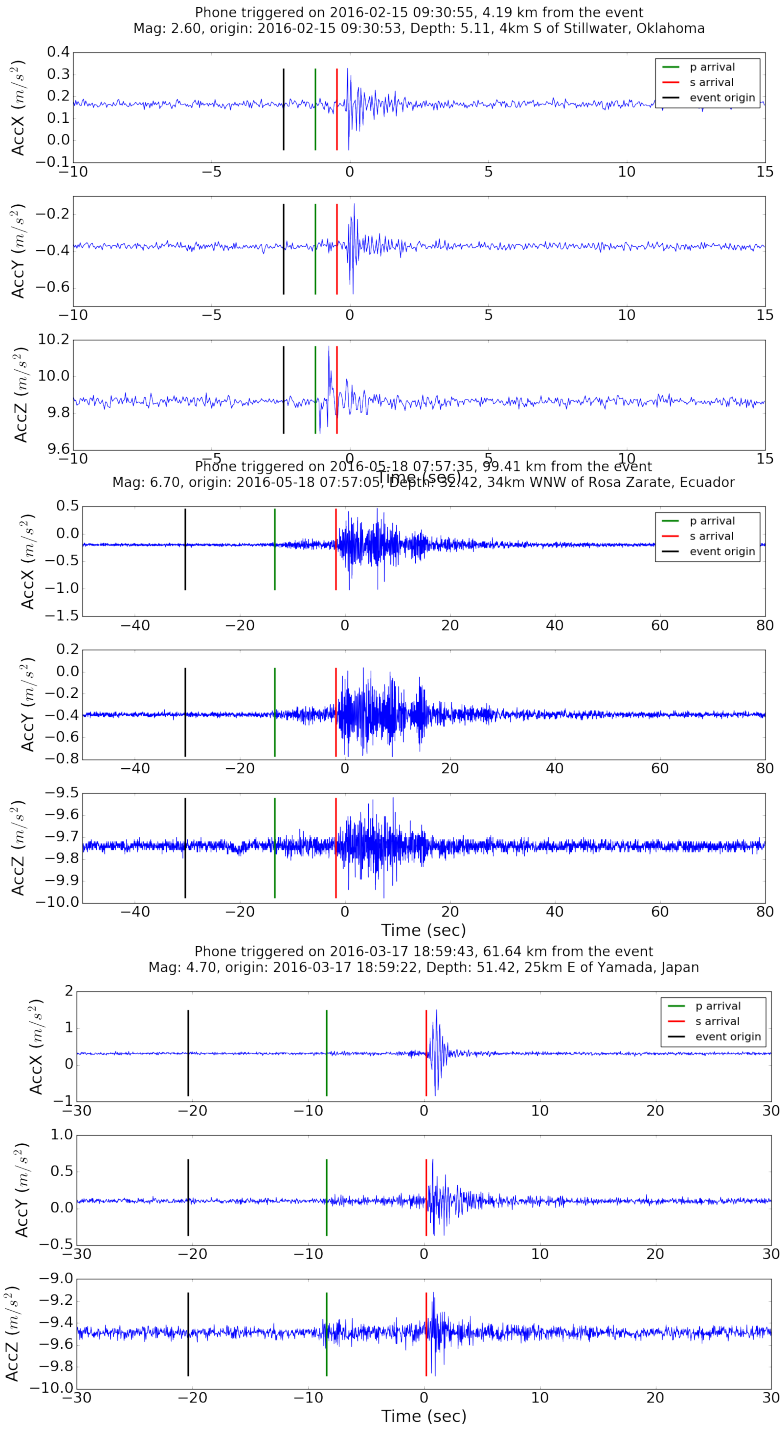
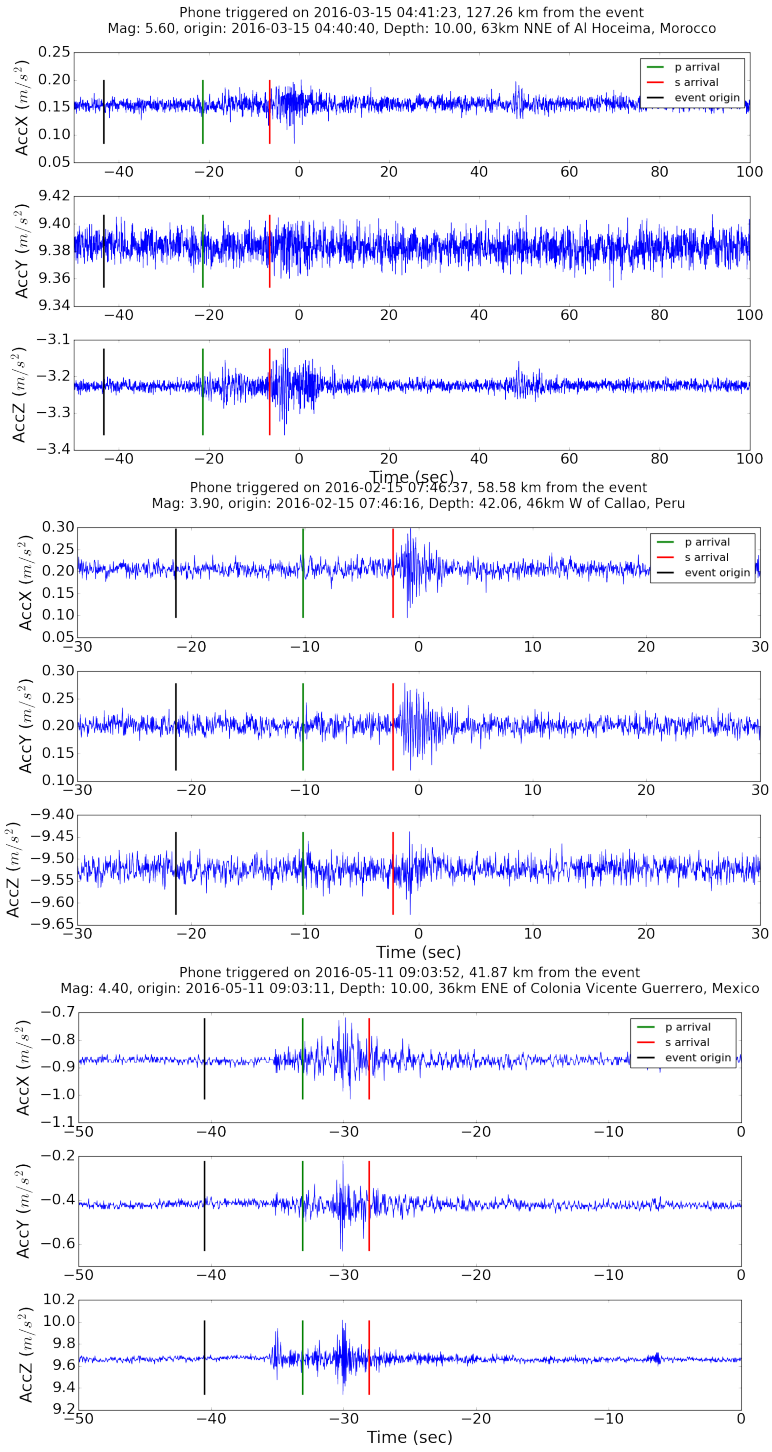


Figure 3.6: Waveform examples from different parts of the world recorded by MyShake. The earthquake location/time is shown. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 min buffer before the trigger. The green and red lines are the estimated P and S arrival times using the ak135 model. The waveform from Nepal has about 1 second missing data, which occasionally happens in MyShake app. Continues below.


Phone triggered on 2016-02-15 09:30:55, 4.19 km from the event
Mag: 2.60, origin: 2016-02-15 09:30:53, Depth: 5.11, 4km S of Stillwater, Oklahoma
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)
Phone triggered on 2016-05-18 07:57:35, 99.41 km from the event
Mag: 6.70, origin: 2016-05-18 07:57:05, Depth: 32.42, 34km WNW of Rosa Zarate, Ecuador
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)
Phone triggered on 2016-03-17 18:59:43, 61.64 km from the event
Mag: 4.70, origin: 2016-03-17 18:59:22, Depth: 51.42, 25km E of Yamada, Japan
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)

Continued


Phone triggered on 2016-03-15 04:41:23, 127.26 km from the event
Mag: 5.60, origin: 2016-03-15 04:40:40, Depth: 10.00, 63km NNE of Al Hoceima, Morocco
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)
Phone triggered on 2016-02-15 07:46:37, 58.58 km from the event
Mag: 3.90, origin: 2016-02-15 07:46:16, Depth: 42.06, 46km W of Callao, Peru
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)
Phone triggered on 2016-05-11 09:03:52, 41.87 km from the event
Mag: 4.40, origin: 2016-05-11 09:03:11, Depth: 10.00, 36km ENE of Colonia Vicente Guerrero, Mexico
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)

Continued

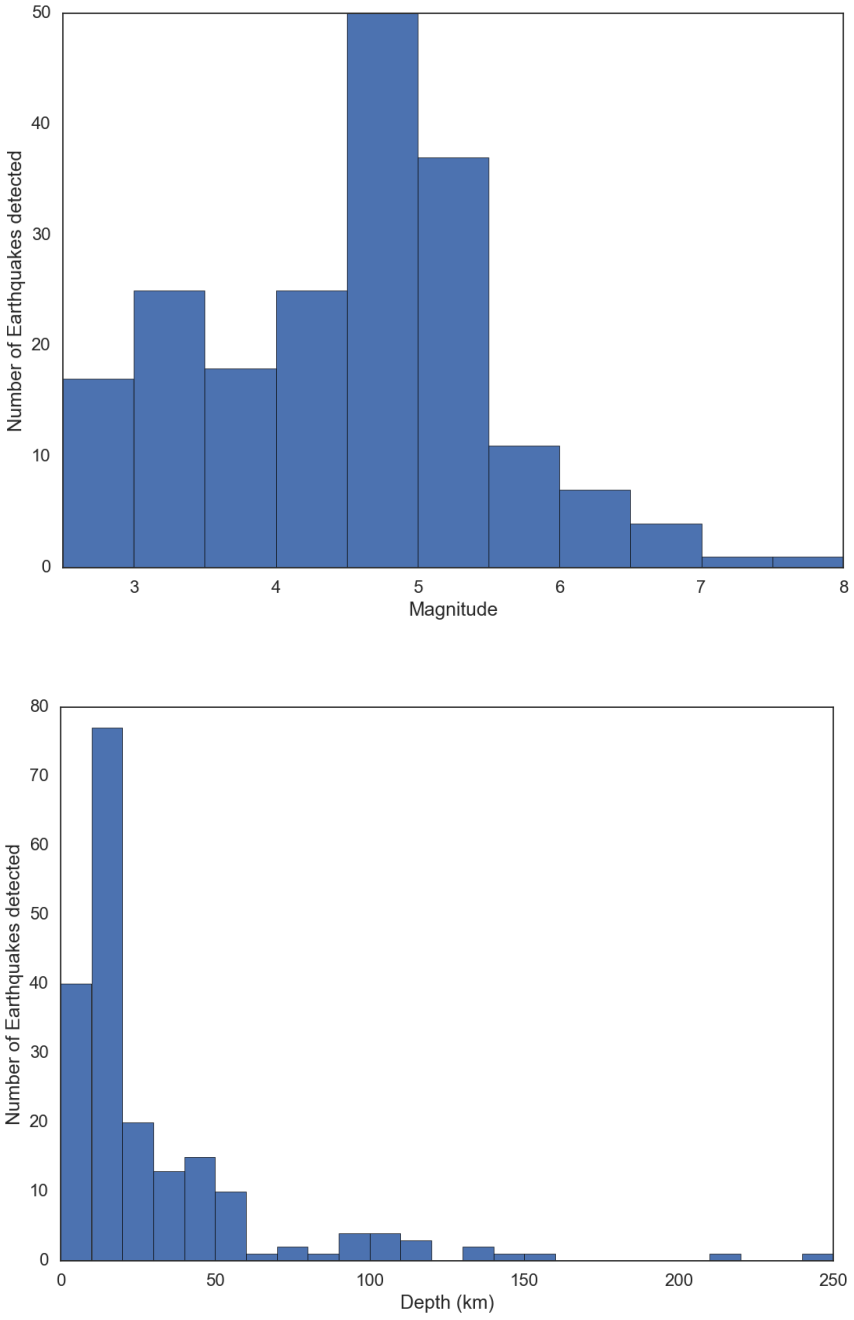


Figure 3.7: Magnitude and depth distribution of MyShake detected earthquakes.

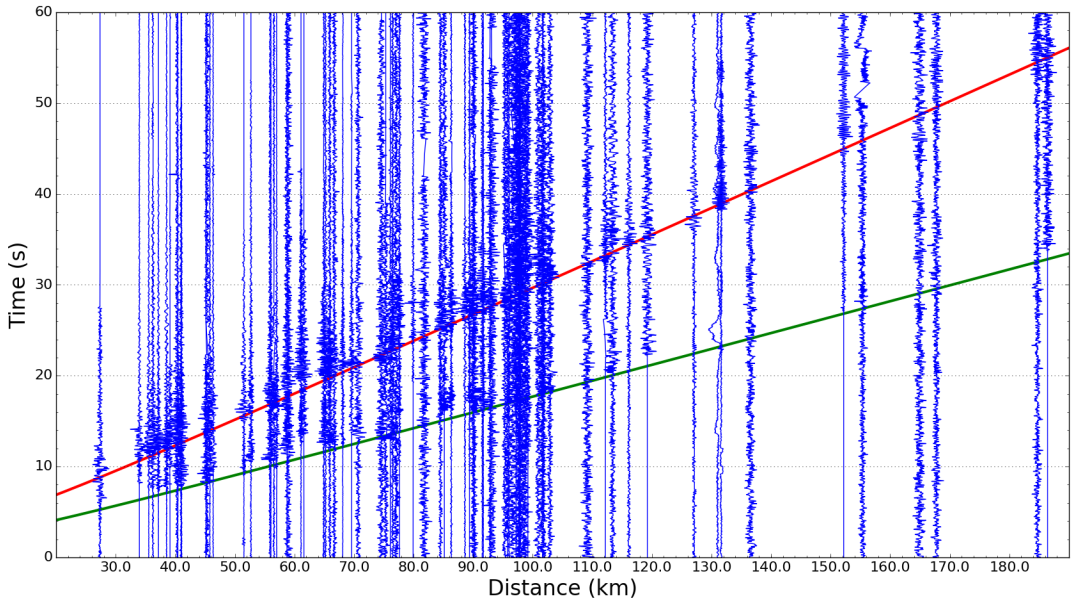


Figure 3.8: Record section plot for the vertical component of MyShake recorded waveforms. The green and red lines are the estimated P- and S-arrival times estimated using ak135.

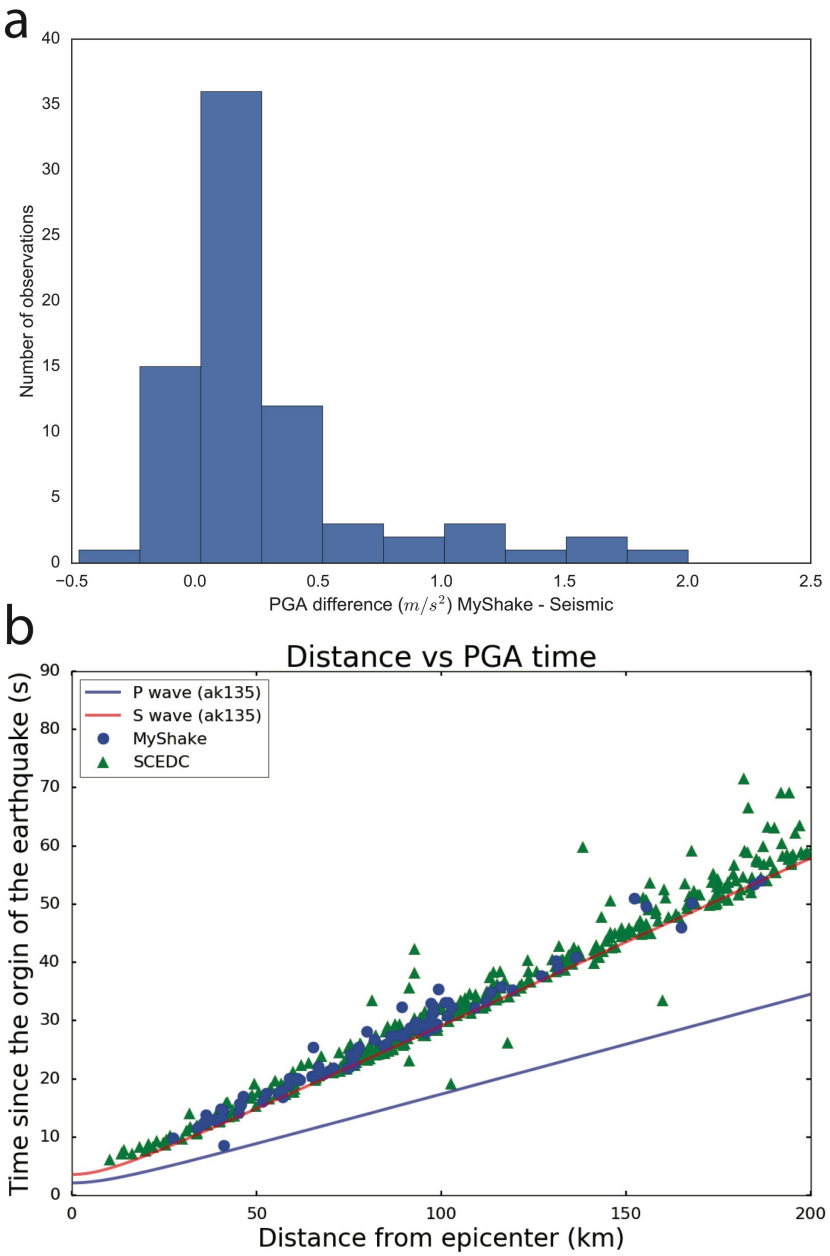


Figure 3.9: PGA comparison and observation times. (a) Histogram of PGA difference between MyShake recordings and the closest traditional seismic station. A positive value means that the PGA value observed by MyShake is larger than that from the traditional seismic station. (b) PGA observation times for MyShake and traditional seismic stations. The blue and red curves are the estimated P- and S-wave arrival using ak135.

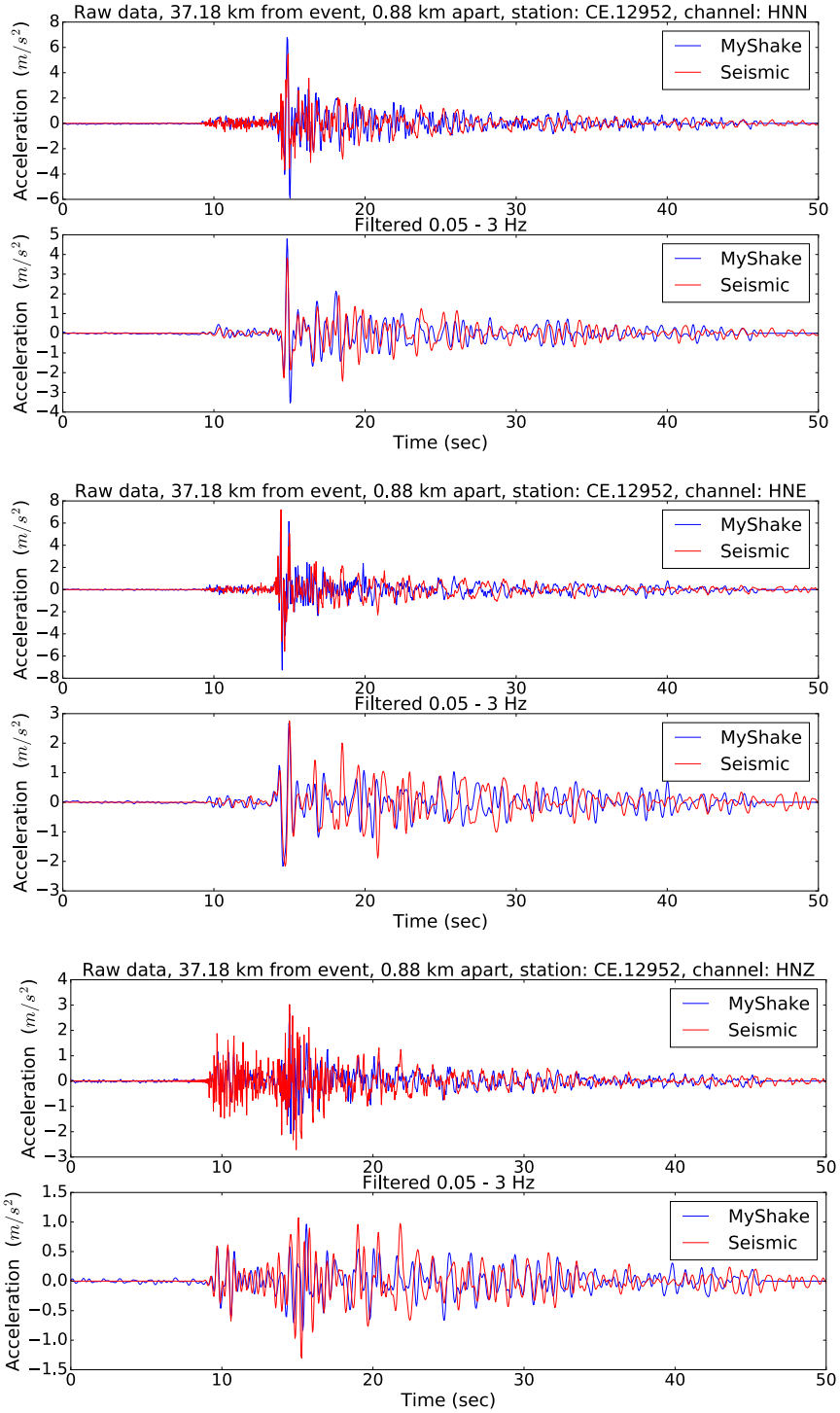


Figure 3.10: Waveform comparisons of all 3 components for a smartphone 37.2 km from the M5.2 Borrego Springs earthquake with a traditional seismic sensor 0.88 km away. The blue traces are recorded by MyShake phone, and the red traces are recorded by the seismic station. Both the raw waveforms and filtered waveforms are shown.

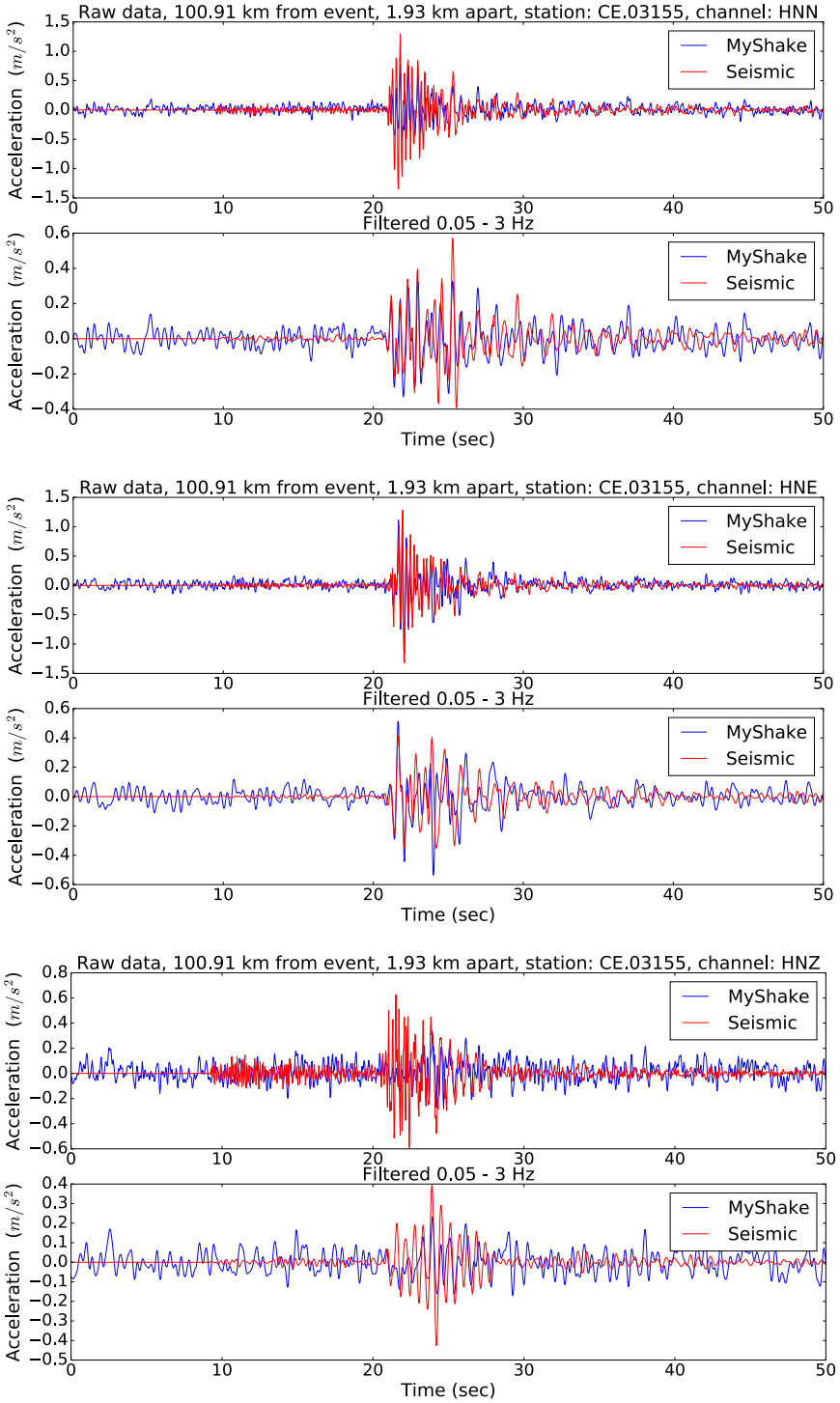


Figure 3.11: Waveform comparisons of all 3 components for a smartphone 100.9 km from the M5.2 Borrego Springs earthquake with a traditional seismic sensor 1.93 km away. The blue traces are recorded by MyShake phone, and the red traces are recorded by the seismic station. Both the raw waveforms and filtered waveforms are shown.

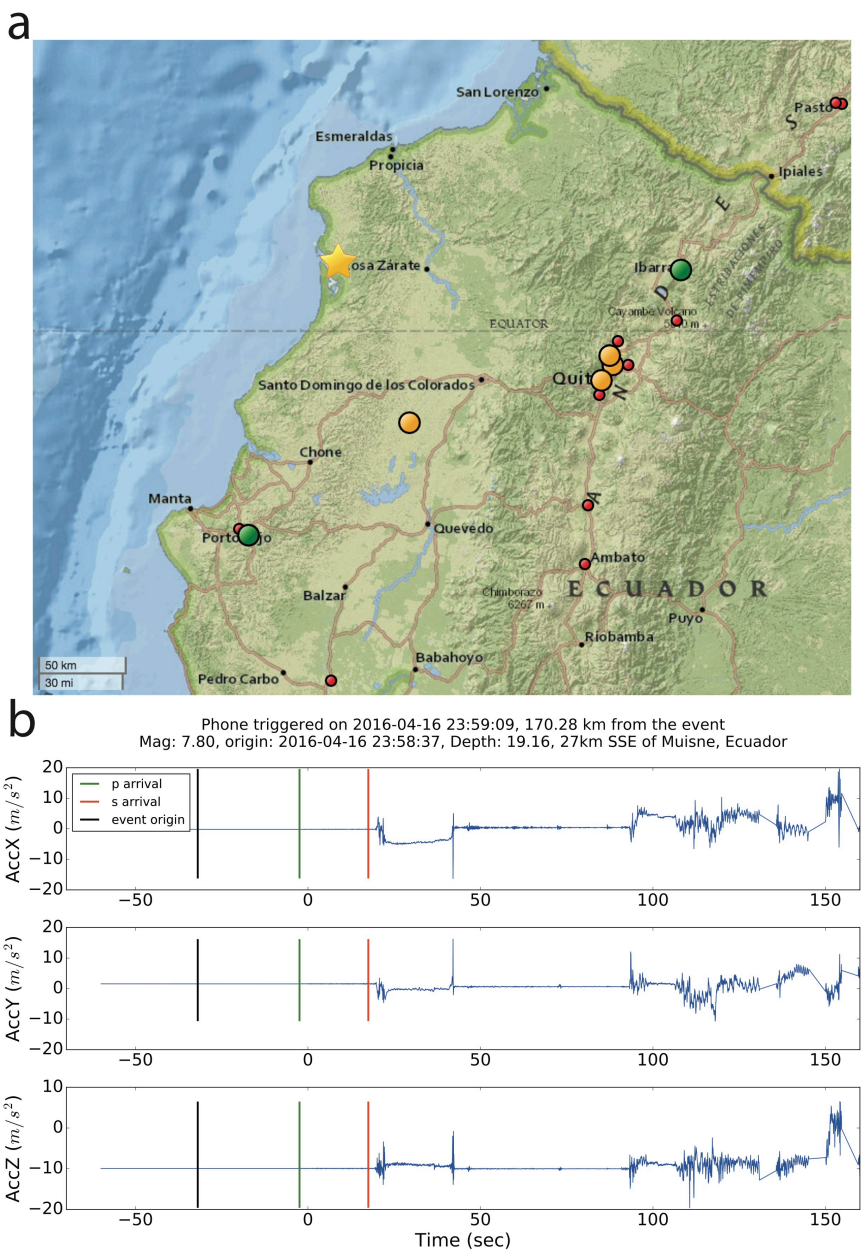


Figure 3.12: Map of the M7.8 Ecuador Earthquake and one 3-component recording. (a) Location of the 2016-04-16 M7.8 Ecuador Earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were ready to detect the earthquakes but did not. (b) 3 component recordings from a user at 170.3 km from the earthquake. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model. The large amplitudes after the S wave arrival are likely from the user picking up the phone when he/she felt the earthquake. The ground motion recorded before the human activities are shown in **Figure 3.4**c.

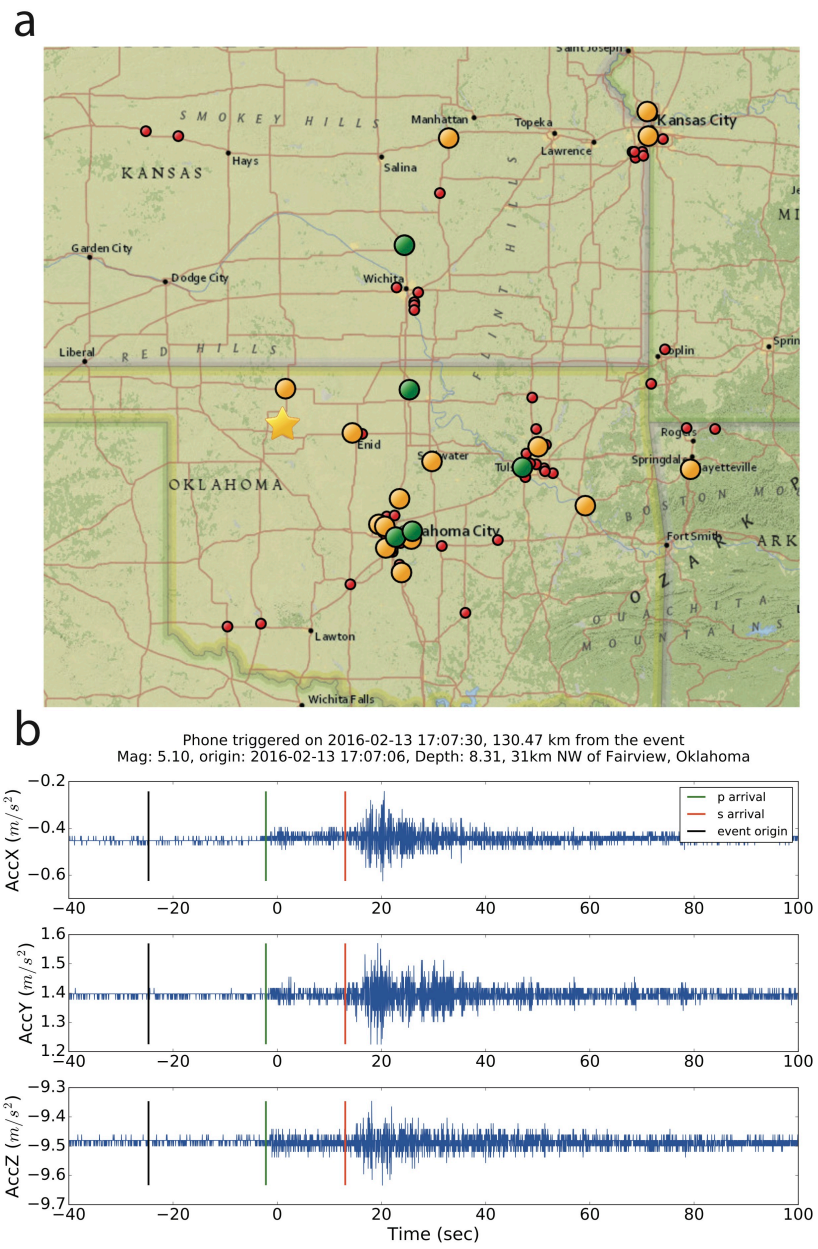


Figure 3.13: Map of the M5.1 Oklahoma earthquake and one 3-component recording. (a) Location of the 2016-02-13 M5.1 Oklahoma Earthquake. Green dots are phones that triggered using the ANN algorithm. The red dots are phones that were not ready to detect earthquakes (likely due to human activities), and the orange dots show the phones that were ready to detect the earthquakes but did not. (b) 3 component recordings from a user at 130.5 km from the earthquake. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model.

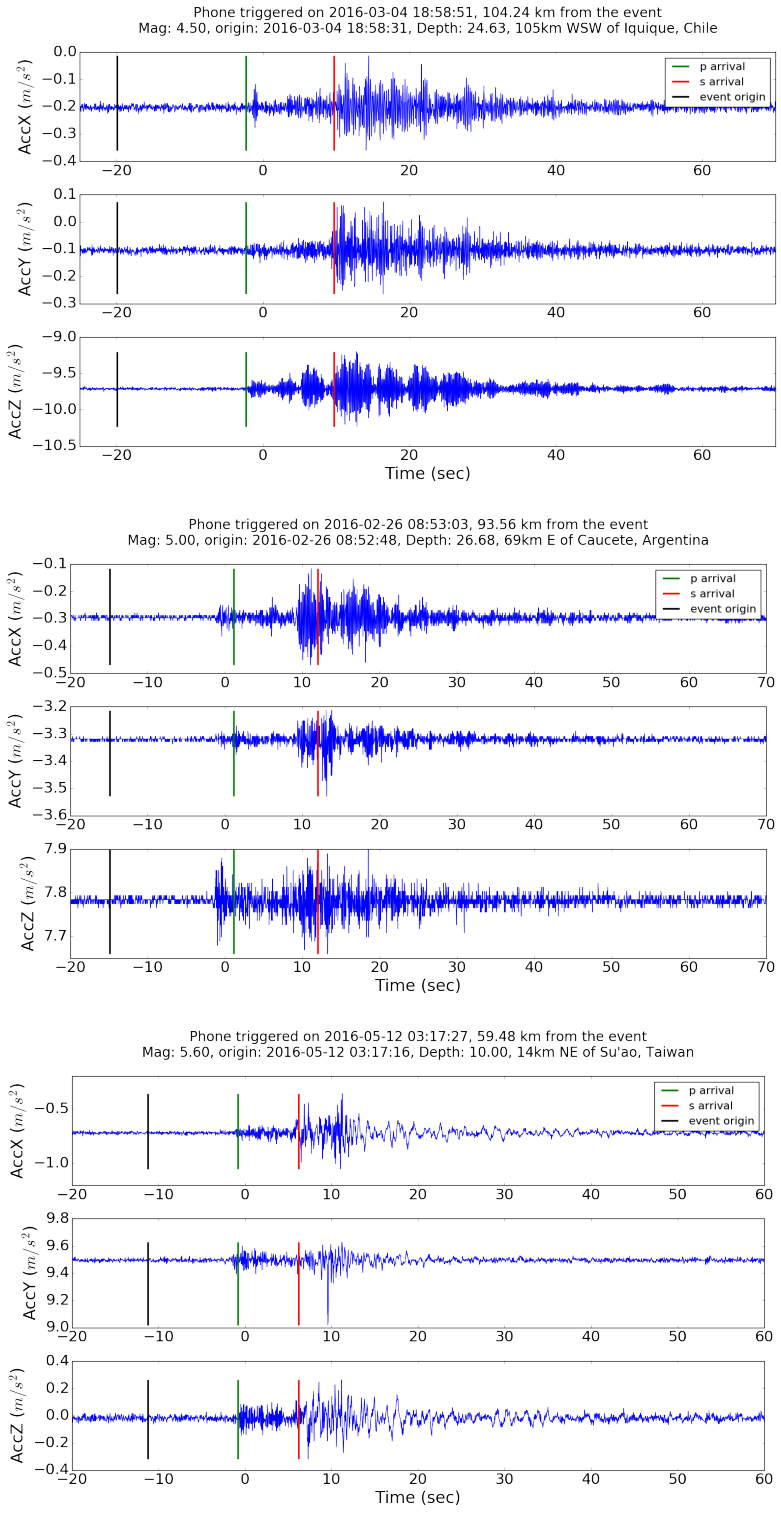


Figure 3.14: Additional examples of P-waves recordings for larger earthquakes. Time zero is the time when the phone triggered, and negative time corresponding the data recorded in the 1 minute buffer before the trigger. The green and red lines are the estimated P and S arrival time estimated by using the ak135 model.

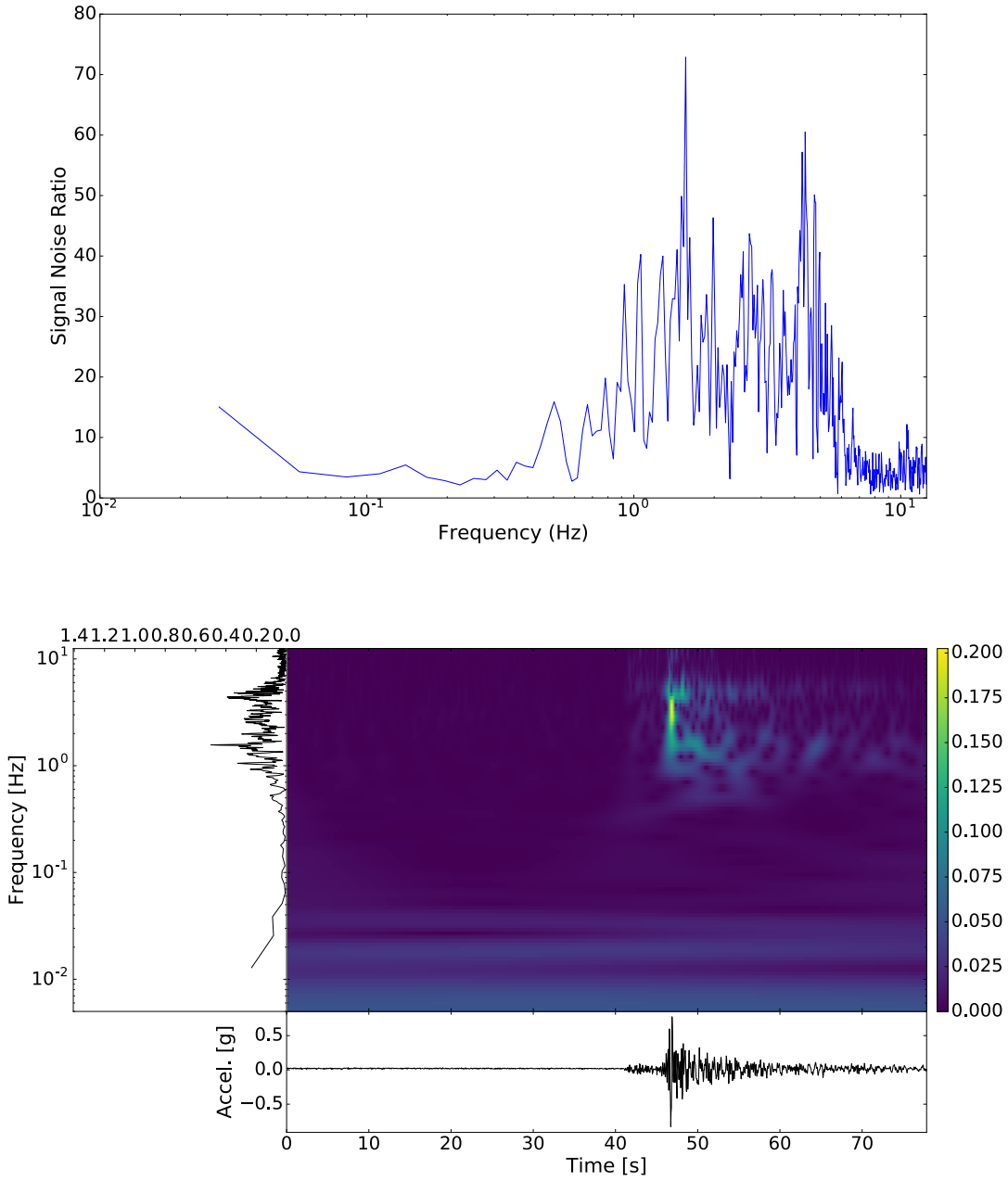


Figure 3.15: Top figure shows an example of the signal-noise-ratio (SNR) as a function of frequency for the horizontal component in **Figure 3.10**. It is computed by using the amplitude of the FFT of the signal divide by the mean noise spectrum amplitude. The phone is at 37.18 km away from the M5.2 earthquake in Southern California. The bottom figure shows time frequency representation of the signal, including frequency spectrum, spectrogram, and the time domain waveform for this recording.

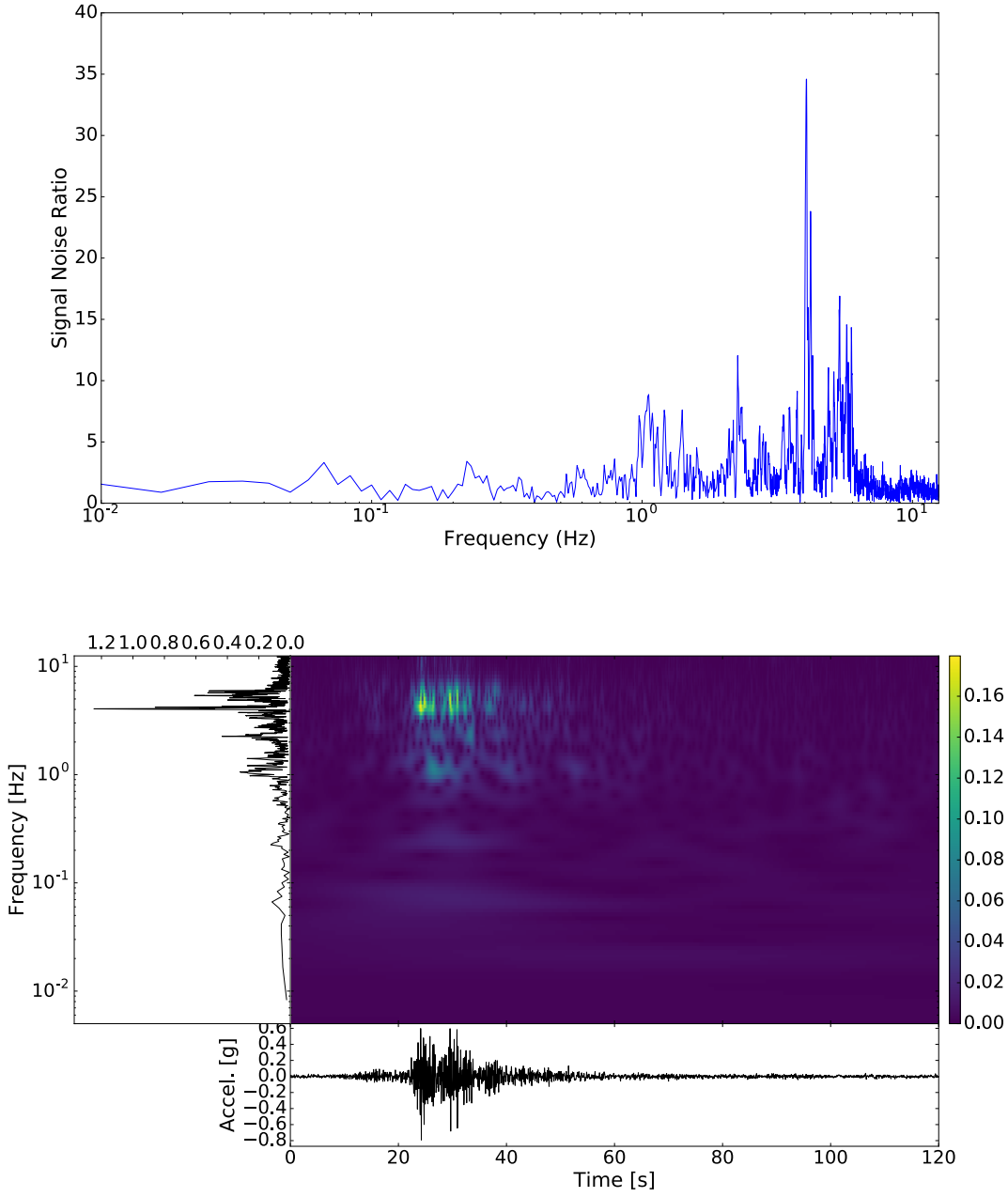


Figure 3.16: Top figure shows an example of the signal-noise-ratio (SNR) as a function of frequency for the horizontal component for the 2016-05-18 07:57:05 UTC, M6.7 Ecuador earthquake. The phone is at 99.41 km away from the earthquake. The bottom figure shows time frequency representation of the signal, including frequency spectrum, spectrogram, and the time domain waveform for this recording.

3.7 Acknowledgments

The MyShake project is a collaboration between Berkeley Seismological Laboratory and Deutsche Telecom Silicone Valley Innovation Center. We thank the MyShake team members: Roman Baumgaertner, Siddartha Pothapragada, Garner Lee, Arno Puder, Pooja Kanchan, Young-Woo Kwon, Stephen Allen, Stephen Thompson, Jennifer Strauss, Doug Neuhauser, Stephane Zuzlewski, and Jennifer Taggart for making this work available. Funding for the project is provided by the Berkeley Seismological Laboratory, Deutsche Telecom, and the Gordon and Betty Moore Foundation through grant GBMF5230 to UC Berkeley. We thank the Northern California Earthquake Data Center (doi:10.7932/NCEDC) and Southern California Earthquake Data Center (http://scedc.caltech.edu/) for providing the seismic data. We also thank Andrew Newman, John Vidale and one anonymous reviewer for helpful reviews. All the analysis of this project is done in python, particularly the ObsPy package [*Beyreuther et al.*, 2010; *Krischer et al.*, 2015]. MyShake project website is http://myshake.berkeley.edu.

Chapter 4 Characterize the earthquakes

4.1 Abstract

MyShake harnesses private/personal smartphones to build a global seismic network. It uses the accelerometers embedded in all smartphones to record ground motions induced by earthquakes, returning recorded waveforms to a central repository for analysis and research. A demonstration of the power of citizen science, MyShake expanded to 6 continents within days of being launched, and has recorded about 800 earthquakes in the first 2 years of operation. The data recorded by MyShake phones has the potential to be used in scientific applications, thereby complimenting current seismic networks. In this paper: (1) we evaluate the capabilities of smartphone sensors to detect earthquakes by analyzing the earthquake waveforms they have collected. (2) We determine the maximum epicentral distance at which MyShake phones can detect earthquakes as a function of magnitude. (3) We then determine the effectiveness of the MyShake network in estimating the location, origin time, and magnitude of earthquakes ("events") which have both P- and S- wave signals above noise levels, and also have more than 5 waveforms for that event. This allows for locations with a mean error of 8.4 km relative to USGS locations for 18 events. For events with only S-wave arrivals available, we were able to estimate the locations as well, although the errors in the locations were larger (with mean error of 14 km) for 41 events. Magnitude estimates can also be determined with mean errors of 0.0 units standard deviation 0.3 units from the USGS catalog, after we corrected the amplitude. Our findings indicate that while MyShake cannot compete with the detection capabilities of traditional seismic networks, these preliminary results suggest that MyShake could provide basic earthquake catalog information in regions that currently have no traditional networks. With an expanding MyShake network, we expect the event detection capabilities to improve and provide useful data on seismicity and hazards.

4.2 Introduction

After more than a century of development, geophysical instrumentation has become more and more diversified. High quality seismic instruments [*Havskov and Alguacil*, 2015], geodetic instruments [*Larson*, 2009], and interferometric synthetic aperture radar [*Bürgmann et al.*, 2000] enable new discoveries and understanding of earthquake physics and active tectonics. In addition, the emergence of various new sensing technologies provides new ways of detecting earthquakes, collecting additional data to learn about the earthquake process, and potentially, making important contributions to seismology [*Allen*, 2012a].

"Did You Feel It," a USGS earthquake survey platform, collects macroseismic intensity data from Internet users which is then used to generate intensity maps immediately following earthquakes [*Survey and Dewey*, 2005; *Atkinson and Wald*, 2007; *Wald et al.*, 2011]. Twitter messages from users who felt an earthquake can be used to detect and characterize events in real-time [*Earle*, 2010; *Earle et al.*, 2010; *Sakaki et al.*, 2010]. By monitoring traffic to its website, the European-Mediterranean Seismological Centre can detect and assess the effect of an earthquake within a few minutes [*Bossu et al.*, 2011]. Low-cost MEMS sensors inside computers

or placed in specially installed stand-alone boxes in homes or offices can be used to monitor and study earthquakes [*Cochran et al.*, 2009b; *Chung et al.*, 2011; *Clayton et al.*, 2011, 2015; *Hsieh et al.*, 2014; *Wu*, 2015; *Wu et al.*, 2016]. Distributed acoustic sensing (DAS) transforms telecommunication fiber-optic cables into seismic arrays, enabling meter-scale recording over kilometers of linear fiber length [*Dou et al.*, 2017; *Lindsey et al.*, 2017].

As more and more people have access to, and a need for smartphones, these small devices comprise an ever-more-widespread and dense network around the globe. Seismologists have learned that smartphones can be used in different ways to detect earthquakes. For example, by monitoring when users turn on a specific earthquake application on their phone, earthquakes can be recognized within minutes as clusters of application activity [*Bossu et al.*, 2015, 2018]. The MEMS sensors inside the smartphones that record acceleration have also been shown to be capable of detecting earthquakes [*Faulkner et al.*, 2011, 2014, *Dashti et al.*, 2012, 2014; *Finazzi*, 2016; *Kong et al.*, 2016a].

MyShake was launched by UC Berkeley in 2016 as a citizen science project. It aims to build a global smartphone seismic network that can be used for research, ultimately contributing to a reduction in earthquake hazards. In the first 2 years, just under 300,000 people downloaded the MyShake app globally. Now, two years after the launch, there are 20,000 phones with the app installed, and on any given day about 8,000 phones contribute data. The core of MyShake is an artificial neural network, built into the on-phone app, that is trained to recognize earthquake-like movement and distinguish it from everyday human movements [*Kong et al.*, 2015, 2016a]. Whenever the phone detects the earthquake-like movement, a 5-minute segment of 3-component acceleration data is stored on the phone and then uploaded to the MyShake server to be analyzed. The time series starts 1-minute before the trigger detected, and continues for 4-minutes post-trigger. Kong et al (2016a) show examples of the waveform data recorded by MyShake, illustrating the potential to use them in different seismological applications. The MyShake data can also potentially be used to monitor the health state of buildings (Kong et al. 2018).

In this paper, we will try to understand the smartphone seismic network by using the data we recorded from daily users. Comparing this data with that of the earthquake catalog, we could see different detection capabilities between the traditional seismic network and this new smartphone network. We will attempt to construct the relationship of expected distance of smartphone network recordings based on magnitude of earthquake. We will also show how, when there are enough recordings, the waveforms recorded by MyShake phones can be used to estimate basic earthquake parameters, including location, origin time and magnitude of the earthquake. This illustrates how the MyShake network could be used to monitor earthquake activity in regions of dense populations that currently have no seismic network. Even given various errors because of the multitude of sources, the results from the MyShake smartphone seismic network look promising.

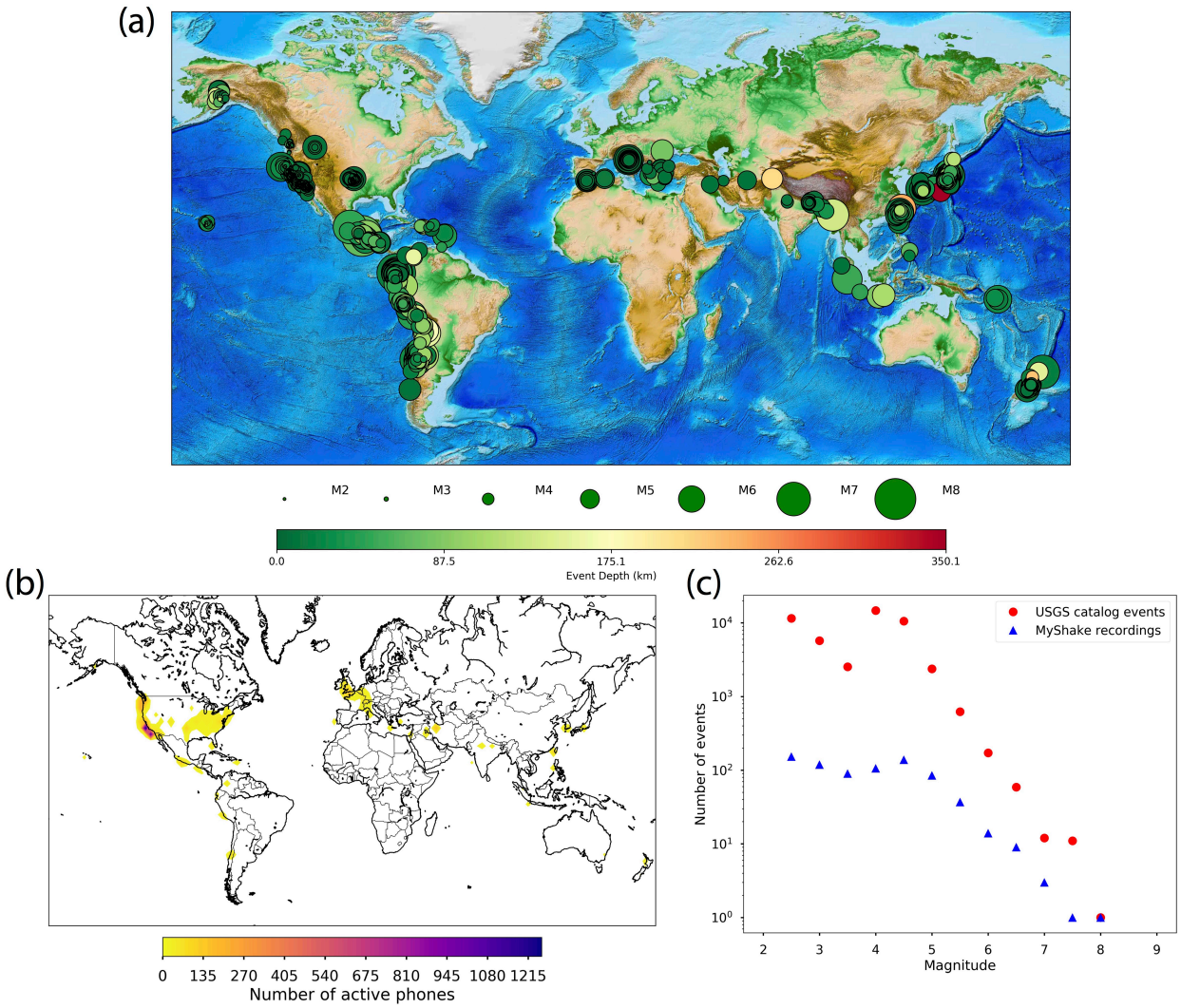


Figure 4.1: (a) Earthquakes with one or more useful waveform recordings from MyShake phones in the first two years of operation (Feb 12, 2016 to Feb 12, 2018). The size of the circle and color represent magnitude and depth of the earthquake (both magnitudes and locations are from USGS ComCat catalog). The small inset figure shows the density of the MyShake users with warmer color showing more users. (b) The heatmap of the number of unique MyShake phones within 3-degree bins. (c) Distribution of earthquake magnitudes for which one or more useful waveforms were recorded.by MyShake (blue triangles) and the catalog events (red circle). The number of events is measured in 0.5 magnitude bins. **Figure 4.11** are showing the difference of the number of events recorded by MyShake and the USGS catalog on the map.

4.3 Data used

The dataset used in this paper comes from global MyShake users. As described in detail in Kong et al. 2016b, the MyShake application has two levels of triggering algorithm, that is

STA/LTA (short-term average and long-term average) [*Allen*, 1978] and the ANN algorithm to determine whether the movement of the phone is due to earthquake or human activities. Once the movements satisfy the ANN algorithm and are determined to be earthquake-like motion, the phone will record a 5-minute segment of 3-component accelerations that will be uploaded to the server later when the phone is connected to WIFI and power. An earthquake waveform database is created from the uploaded waveforms by scanning for "candidate events" registered in publically available USGS (United States Geological Survey) ComCat catalogs. For each "candidate event" from the USGS catalog, we search within a pre-defined spatiotemporal window for possible triggers in our smartphone waveforms database. Waveforms that meet the requirements of the above spatiotemporal window are reviewed by a seismologist to filter out those caused by human activities, and to remove the so-called "incomplete waveforms," which are missing blocks of data. Waveforms that pass all the checks are put into the earthquake database. In the first two years, 757 earthquakes have had at least one recording from a MyShake user.

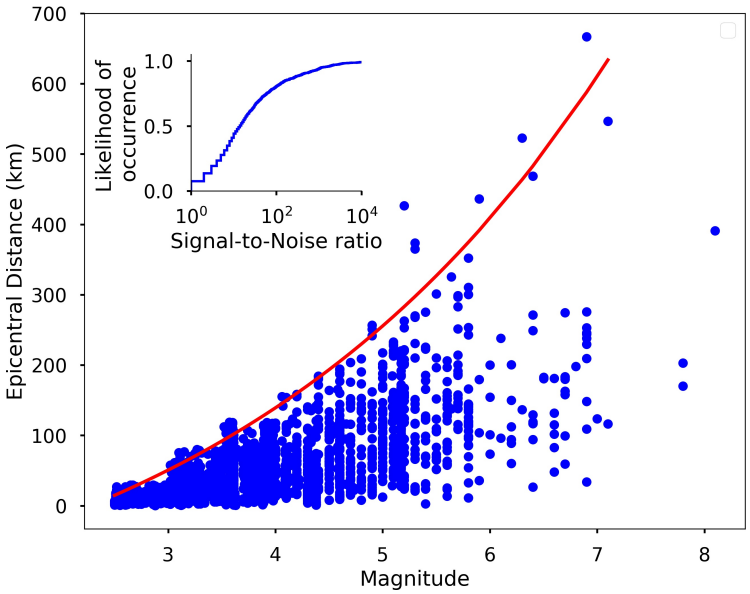


Figure 4.2: Distance of waveform recordings recorded by MyShake as a function of magnitude. The blue dots are waveforms recorded by MyShake, and the red curve is the fitted line to the furthest recordings from MyShake database. We only search M2.5 and above events that are corresponding to the USGS catalog event using a space and time window. The red curve is the analytic representation for estimate the furthest waveforms we expect to see for different magnitudes. The inset figure on the top left is the cumulative distribution of the signal to noise ratio (SNR) for all the earthquake recordings measured on one horizontal component (Y component).

Figure 4.1a shows the location of earthquakes for which one or more seismic waveforms (confirmed by a seismologist) was uploaded from MyShake phones. The corresponding unique MyShake user distributions are shown in Figure 4.1b. We can see that with more phones, more

events have been recorded. Figure 4.1c shows the number of events recorded by MyShake that are within 0.5 magnitude of USGS ComCat catalog events. For almost all the magnitude bins, MyShake records fewer events than the traditional seismic network, but for magnitude 4.5 and above, MyShake follows a slightly different slope line, which indicates that MyShake has better detection capability for larger magnitude earthquakes. At smaller than M4.5, we see that the gap between MyShake and the catalog increases as magnitude decreases.

Waveforms uploaded to our server are 3-component acceleration waveforms, in 5-minute segments (1-min before the earthquake and 4-min after), sampled at about 25 Hz. Figure 4.2 shows epicentral distances of the earthquake waveforms recorded by MyShake users for earthquakes of various magnitudes. Obviously, as the magnitude of the earthquakes increases, the distance from which smartphones can record useful waveforms also increases. To understand at what range we expect MyShake users to record earthquake waveforms, we fit an analytic expression to the furthest recordings in each magnitude bin using least-square regression. For earthquakes of M2.5 to M8.0, we derived the following relationship between the magnitude of the earthquake and the distance in kilometers from which we expect to see recordings from current MyShake users:

$$Dist. = 122.860e^{0.275mag} - 229.348$$

The “mag” and “Dist” are magnitude and epicentral distance for the earthquake. The curve of the above equation is shown as the red curve in Figure 4.2. The small inset figure in Figure 4.2 shows the cumulative distribution of the signal to noise ratio (SNR) for all the earthquake waveforms from MyShake users. The SNR is calculated on the Y-component by selecting a 2-second signal from the PGA value (±1 sec around it), and a 2-second noise from the beginning of the waveform (we have 1 min noise before the trigger). Within the 2-second window for the signal and noise, we calculate the square of the root mean square (RMS) amplitude and take the ratio. The 25, 50, and 75 percentile of the SNR are 6.1, 14.6 and 50.9 for the MyShake recorded earthquake waveforms.

Figure 4.3 shows six selected waveforms recorded around the globe by MyShake users. We focused this selection on locations where we don’t have dense seismic networks; additional waveforms are available in the supplementary material. The X and Y components of the acceleration records are parallel to the short and long direction of the phone-screen plane, while the Z component is the direction perpendicular to the phone screen. On most of these waveforms, we can see clear P- and S-waves. Because we also record 1-min of data before the phone triggers, even when the phone triggers on the S phase we still have the P-wave recorded if it is above the noise.

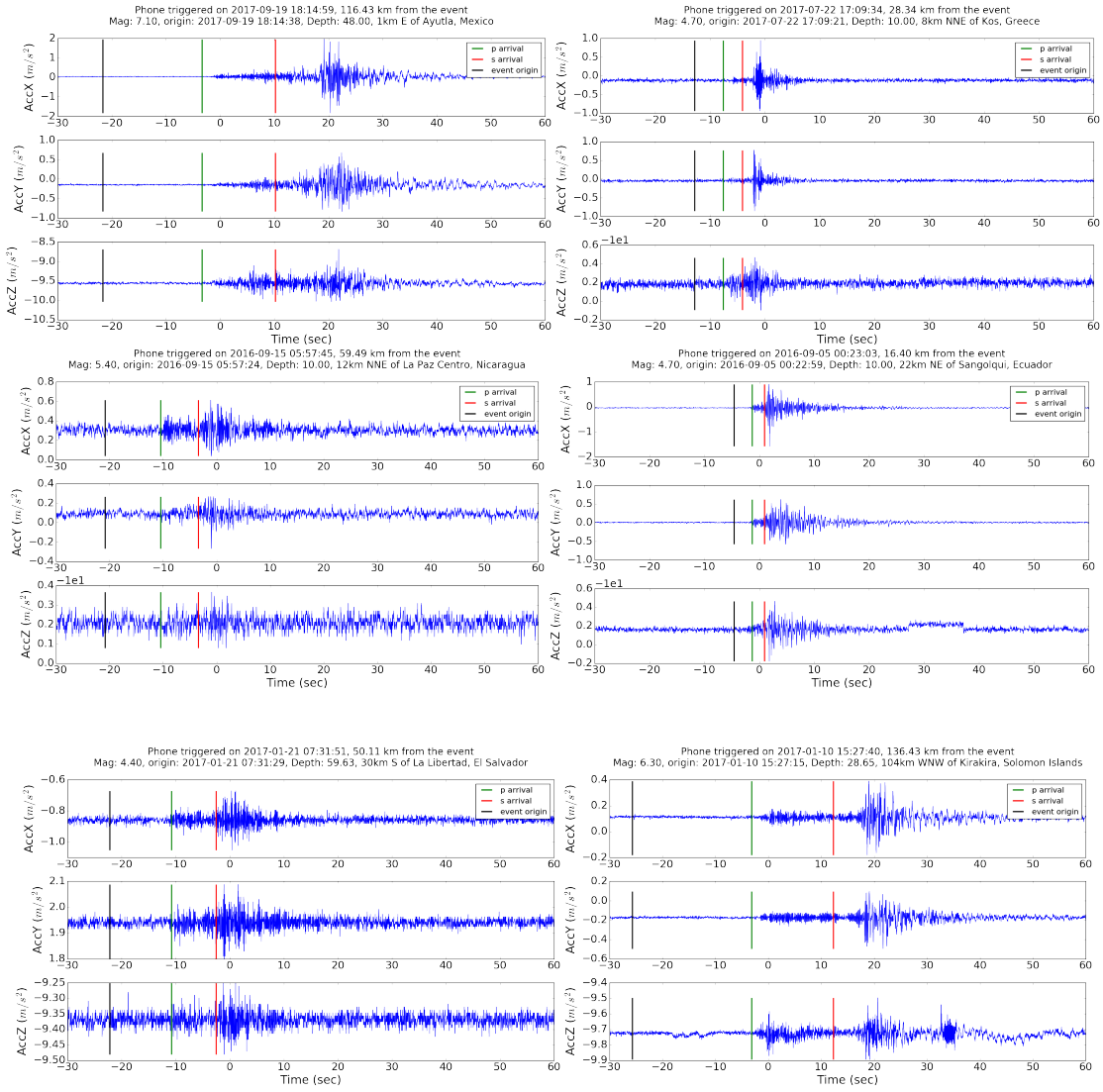


Figure 4.3: 3-component acceleration waveforms from MyShake detections globally. The black line is the event origin time from USGS catalog, green and red lines are estimated P and S arrival time from model ak135 [*Kennett et al.*, 1995]. Time 0 on each panel is the time when the phone thinks it detects an earthquake.

4.4 Timing and location accuracy of smartphone records

Since the location and time service on the smartphones will contain errors, we try to estimate and adjust for errors in the data we have collected. In this session, we report the location error and timing error from our experiment and the data we collected from the MyShake users.

MyShake requests GPS locations from participating phones, which is available only in cases where users have agreed to share that information via their GPS location service. In order to get a better understanding of typical "location errors" for smartphone GPS systems, we did a test by placing 20 smartphones near a window in a 9-story building on the 9th floor for 12 hours, and sampled the GPS location at about 25 Hz. This is the ideal case: the phones are stationary,

and are placed near a window, inside a building. In Figure 4.12, we plot the mean location error and the standard deviation from the phone measurements to a reference point near the phone, with latitude and longitude obtained from Google maps. Overall, we could see the errors on the phones are very small, with standard deviation all within 0.4 m. There was only one phone (id 6) that had a mean distance error of 557 m with standard deviation 0.001 m, for unknown reasons. Typically smartphone GPS locations are accurate to within a 4.9 m radius under open sky, however, their accuracy worsens near buildings, bridges, and trees (from Official U.S. government information about the Global Positioning System).

As the smartphone clock could drift over time, a Network Time Protocol (NTP) in MyShake corrects the internal clock of the phones on an hourly basis. To facilitate an assessment of the accuracy of phone timing, we retained the data and statistics returned by each phone as part of the NTP. To represent the overall trend, we sampled and analyzed 24 random day-long 'snapshots' from all MyShake users, for a twelve-month span beginning in August of 2016 These 24 days encompassed ~6.2 million usable records, each of which represents a successful NTP query by a phone running MyShake. The statistics collected included the round trip time for the phone's NTP query to reach and return from the server, as well as the offset in milliseconds of the server time relative to that of the phone's internal time when the query was initiated. The total actual offset ('true offset') was calculated as the recorded offset minus the one-way time lag—approximated as half the round trip time—required for the query to reach the server. The round trip time is typically an order of magnitude smaller than the recorded offset (over 60% the reported values are < .01 seconds). Figure 4.13 shows the boxplot for the corrected offsets from this dataset, which represents the distribution of the phone timing uncertainty within 1 hour. For current MyShake users, 75% of the timing offsets are within 1.6 s, and 85% of the timing offsets are within 4.9 s. The median offset is 0.8 s, but at 95 percentile, the offsets go up to about 88 s.

From the location and timing error estimation, we could see that the errors from the locations are usually small, and the timing errors are mostly small within 2s, but could be up to tens of seconds on some phones. These waveforms with very large timing offsets could be identified on the server, since when the phone triggers it sends a trigger message with the time on the phone to the server. By comparing this trigger message with the time it arrives, a simple filter could be set up to remove these large offsets triggers and waveforms.

4.5 Location and origin time estimation

One goal for MyShake is to detect earthquake and characterize the earthquake parameter to obtain reasonable estimates of the location, origin time, and magnitude of detected earthquakes. We selected earthquakes that have more than 5 waveforms, which are a total of 69 earthquakes globally from February 12, 2016 to February 12, 2018, as shown in Figure 4.4. We manually selected all of the P phases and S phases for all 69 events. Due to signal to noise ratio, not all the waveforms will have a P phase or S phase picked out. Therefore, we ended up with 18 events with more than 5 waveforms, and with both P and S wave arrivals selected, and 41 events with only S wave arrivals selected.

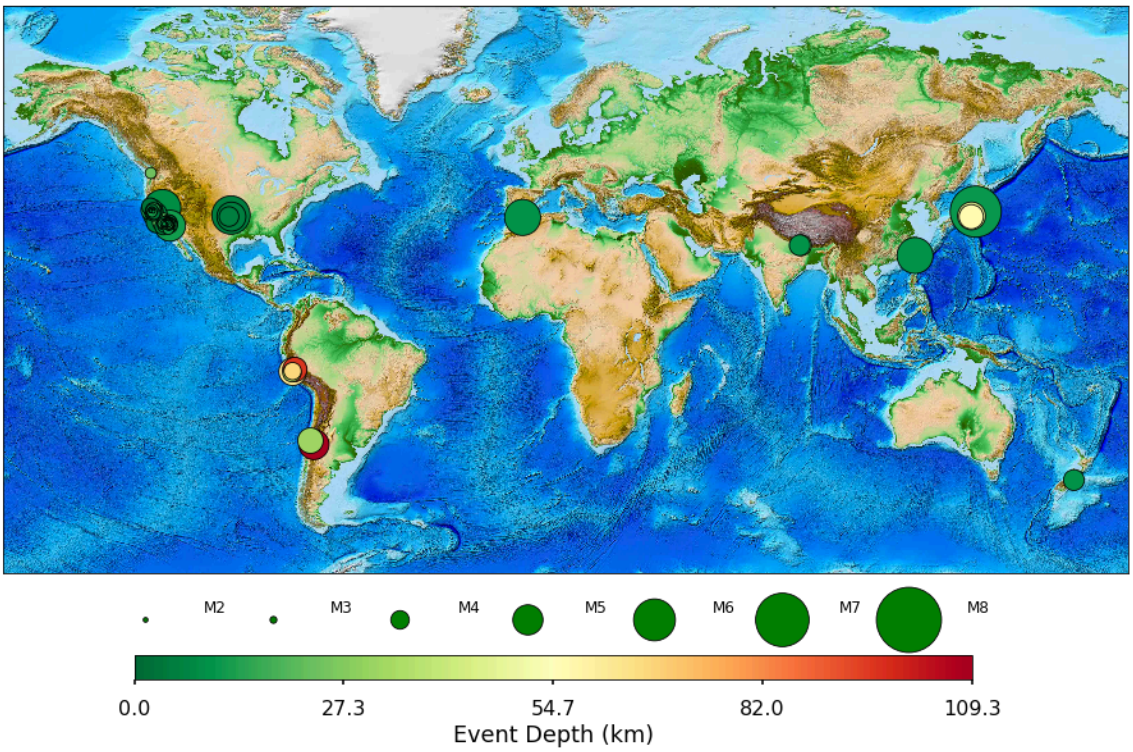


Figure 4.4: The location of the 69 selected events with more than 5 useful waveforms for anlysis. Size of the circle represents the magnitude and color coded by the depth of the event.

If both P and S phase selections are available, to estimate the location of the earthquake, we could do a grid-search based on the time difference between the P and S waves. We minimize the squared error between the observed P, S difference and the estimated ones on each waveform. The search grid extends ±2 degrees in latitude and longitude around the centroid of all the available phones and we search on a grid with 0.2-degree step. After finding the best solution, we then start a finer grid around this solution by extending ±0.2 degrees in latitude and longitude with a 0.01-degree step. The depths of all these events are currently fixed at 8 km. For the travel time estimates, we used a P-velocity of 6.10 km/s and an S-velocity of 3.55 km/s and we assumed straight rays. We call this "ElarmS approach," and it has been shown to be a reasonable approach and approximation for real-time detection systems in California [*Allen et al.*, 2009a; *Kuyuk et al.*, 2014]. (We also tested using the ak135 velocity model and got very similar results. Therefore, for simplicity and faster computation, we choose this simple velocity model for the current dataset.) The result of the grid search is shown in Figure 4.5a. We can see that the mean epicentral distance error for these events is 8.4 km with a standard deviation of 7.6 km. A comparison of these event locations, estimated by using only the S phase selections, is shown in Figure 4.5b (the method is described in the next two paragraphs), which has epicentral distance error 14.0 km with 19.0 km standard deviation. With both P and S phase selections, we could have better results than with just S phase selections.

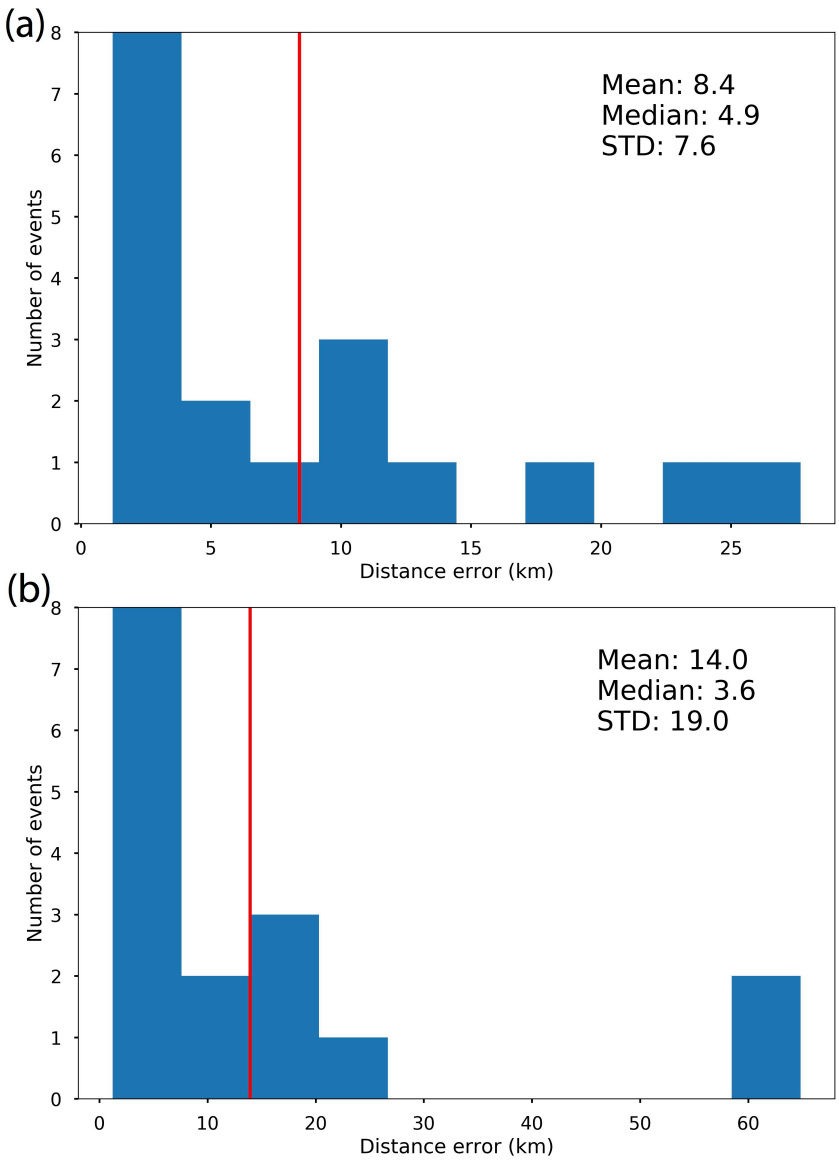


Figure 4.5: (a) The distribution of the epicentral distance errors from the estimation using the time difference between P and S phases. The red line is showing the mean value. There are total of 18 events. (b) The distribution of the corresponding 18 events epicentral distance errors from the estimation using S phase pickings alone. The red line is showing the mean value.

Our objective is to use MyShake to detect and locate as many earthquakes as possible. For this reason, we are assessing the ability of using just the S-wave to locate events. The relatively high noise floor of the MEMS (Micro Electro Mechanical Systems) sensor built into smartphones [*Kong et al.*, 2016a] means that the P-wave arrival is not always visible on the waveforms, especially for small earthquakes or phones at large distances. However, the S-wave arrivals are clear on the majority of waveforms recorded by MyShake users, as described above, where we have 41 events having more than 5 waveforms and a defined S phase.

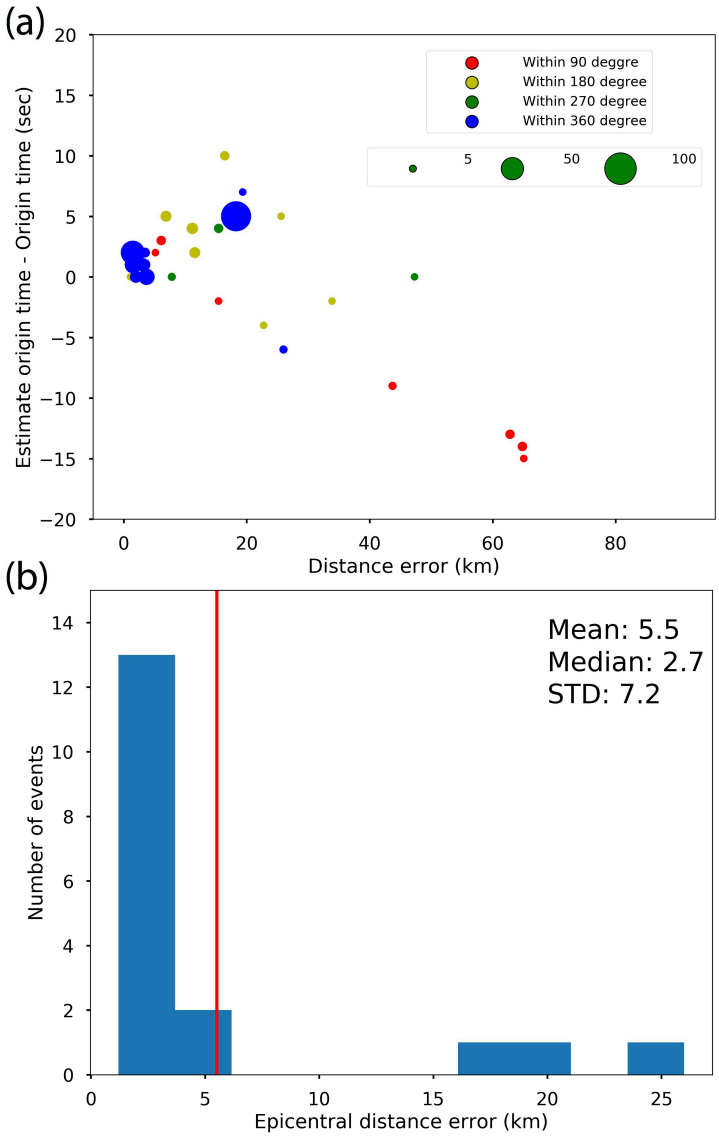


Figure 4.6: (a) Estimated origin time error vs the epicentral distance error using only the S phase pickings. The size of the circle is showing the number of waveforms used in the calculation and the color is showing the station coverages that estimated from the azimuth of the two furthest stations and check whether it is in the range of 90 degree, 180 degree, 270 degree, and 360 degree as an indicator of the station coverage. The mean epicentral distance error is 17.8 km, median error 6.1 km and with a standard deviation 22.8 km. The mean origin time residual is 0.1 s (Estimate origin time – Catalog origin time), median error is 1.0 s, and standard deviation 6.9 s. (b) The histogram of the epicentral distance error from only the best coveraged events which have the phones' azimuthal range larger than 270 degree. The red line is the mean value 5.5 km, the median is 2.7 km and standard deviation is 7.2 km.

We next use the grid-search method to estimate the latitude, longitude and origin time of the earthquake by minimizing the squared error of the estimated and the observed S-wave arrival time on each waveform. We currently fixed the depth at 8 km in the grid-search. For the travel

time estimates, we again use both the ak135 velocity model and the simplified ElarmS approach and get very similar results, therefore, we only report the one using the ElarmS approach. The search grid extends ±2 degrees in latitude and longitude around the centroid of all the available stations and we search on a grid with 0.2-degree step. The origin time search starts from the earliest station trigger time to 60 s before it in 1 sec steps. Once we find the best solution from this grid-search, we will have a finer grid to search based on the best solution from the coarse grid-search. The finer grid extends ±0.2 degrees from the coarse solution with a step at 0.01-degree, while the origin time search extends ±10 s with 1 s as step.

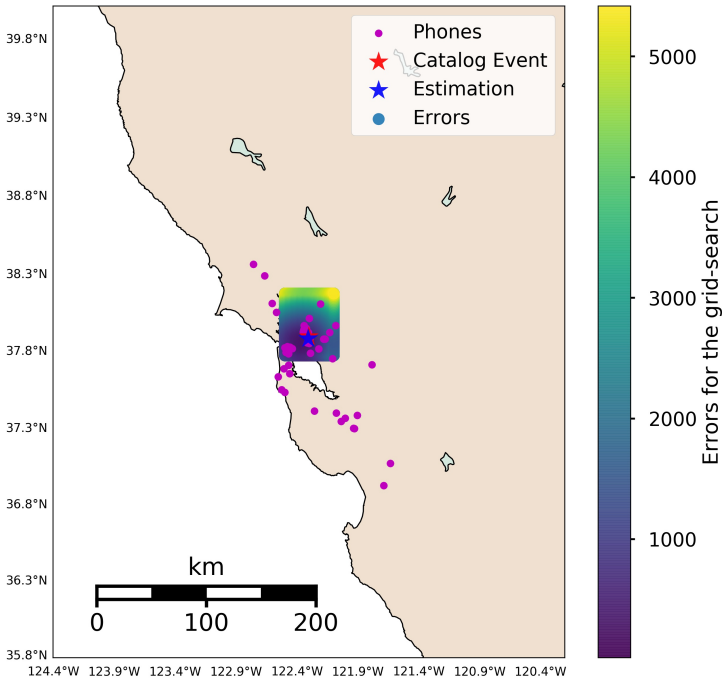


Figure 4.7: Location estimation example from a well-recorded event - M4.4 event in Berkeley on 2018-01-04, 10:39:37.730 (UTC). The magenta dots are the location of the waveforms used in the analysis. The red star is the USGS catalog location and the blue star is the estimation location. The error in the epicentral distance is 1.5 km and with a 2 s origin time error. The grid-search is starting from the whole region showing in the map and finds the best solution at a coarse grid. The grid dots showing here are the finer grids in the analysis that color-coded by the sum of the square of the difference between the observed S phase time and the estimated S phase.

The results of the final finer grid-search locations and origin times are shown in Figure 4.6a. The mean epicentral distance error is 17.8 km, median error 6.1 km and with a standard deviation 22.8 km. The mean origin time residual is 0.1 s (Estimate origin time – Catalog origin time), median error is 1.0 s, and standard deviation 6.9 s. The circles in the figure are sized by the number of waveforms used in the calculation and colored by the phone coverage. In general, the events with good azimuthal coverage of the phones and the more waveforms tend to have better location and origin time estimation. All the events except 1 with an error more than 40 km have only very limited station coverage and the event occur outside of the network.

Therefore, we only pick out the events with good phone azimuthal coverage – above 270 degree (the ones showing as blue dots) and compute the location errors, which is shown in Figure 4.6b. The mean distance error is 5.5 km, the median is 2.7 km and standard deviation is 7.2 km, which are significantly better compare with the events without good coverage.

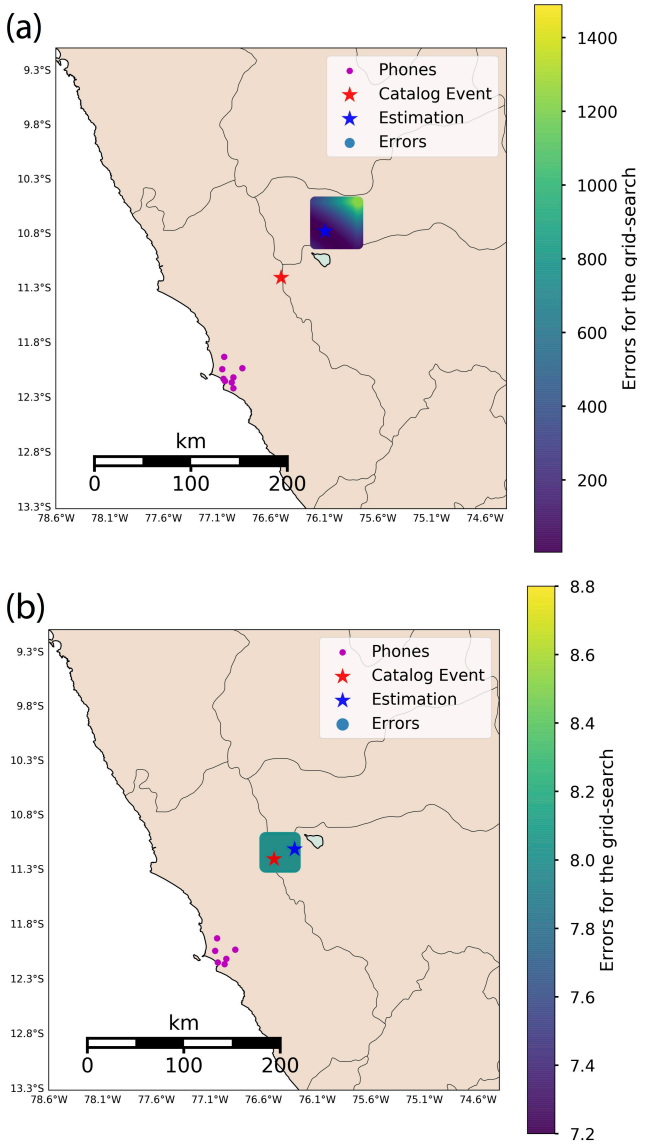


Figure 4.8: (a) Location estimation example from M4.6 event in Peru on 2017-02-01, 11:38:30.100 (UTC). The magenta dots are the location of the waveforms used in the analysis. The red star is the USGS catalog location and the blue star is the estimation location from the finer grid search using only S phase pickings. The epicentral distance error is 64.8 km between the estimated and catalog location. The grid-search is starting from the whole region showing in the map and finds the best solution at a coarse grid. The grid dots showing here are the finer grid in the analysis that color-coded by the sum of the square of the difference between the observed S phase time and the estimated S phase (b) Location estimation for the same events but using both P and S phase pickings. The location error is 23.2 km.

Figure 4.7 shows an example of our estimation at Berkeley, CA, USA where we have good phone coverage. With the finer grid-search, the estimated location error and origin time error are 1.5 km and 2 s from the corresponding catalog counterparts. Figure 4.8a shows an example of our estimation at a place where there is not a good traditional seismic network – Peru. We only picked 8 S-wave phases from the MyShake users (shown as the magenta dots in the figure), the estimated location and origin time errors are 64.8 km and -14 s from the corresponding catalog counterparts. Given that this event was outside of the phone network, and the azimuthal coverage of the MyShake phones is only 12.6 degrees, this is a reasonable location. As for this event, we can pick both P and S wave arrivals for more than 5 phones, therefore, we can minimize the relative time between P and S wave. The location will be improved because we have more information to constrain the results. Figure 4.8b shows the results of the location estimation error significantly improved from 64.8 km to 23.2 km after taking into account the P phase selections as well. Therefore, for earthquakes where we can extract both P and S wave arrivals, our estimation could be better than when using S waves alone.

4.6 Magnitude estimation

The magnitude is another parameter to characterize the earthquake. It is important to obtain a reasonable magnitude of an earthquake from MyShake recorded data. Currently, we have many events from California as we have the densest concentration of MyShake users here. Therefore, we apply the relationship of estimating the ML magnitude in California [*Bakun and Joyner*, 1984] as our starting point to estimate the magnitude.

With more data to be recorded in the future from various locations, we can fine-tune relationships for different regions around the globe. For estimating the ML magnitude, we use the "estimate magnitude" function from ObsPy [*Beyreuther et al.*, 2010; *Megies et al.*, 2011], which estimates local magnitude from the distance, poles and zeros, the peak to peak amplitude, and the time span from peak to peak.

In Figure 4.9a, the estimated magnitudes using the estimated locations from above are plotted against the USGS catalog magnitudes. Most of the magnitudes estimations are within 1 magnitude unit difference, with the mean difference 0.3 and standard deviation 0.4 (estimated – catalog magnitude). Only 4 events have a magnitude estimation difference slightly larger than 1 unit, and all of them except 1 has relatively fewer waveforms. Another observation is that most of the magnitude estimations are higher than the USGS catalog magnitudes. We can see the histogram in the inset is skewed to the left. We discussed in the initial observations [*Kong et al.*, 2016b] that MyShake recordings usually have larger amplitude than the nearby seismic stations, which is likely due to the fact that the phones are not coupled to the ground, but instead placed on the desk or places in the buildings. Therefore, these larger magnitude estimations are likely due to this amplitude difference.

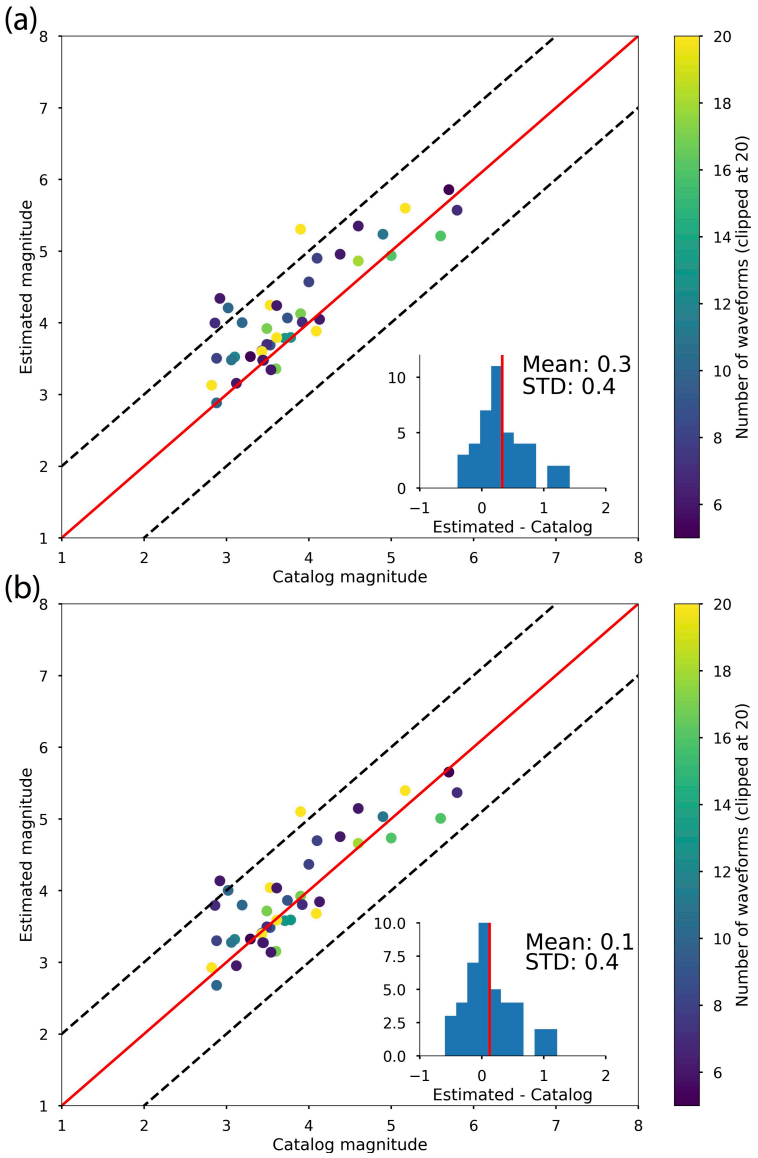


Figure 4.9: (a) Estimated magnitudes versus USGS catalog magnitudes using the estimated locations. The red line is 1 to 1 magnitude, and the two dashed lines are 1 magnitude unit off from the 1 to 1 line. Each dot is estimated using different number of waveforms illustrated by the color. The inset plot is histogram of the residual magnitude (estimated magnitude – catalog magnitude). The mean of the residual is 0.3, and standard deviation is 0.4. The line in the histogram is the mean value. (b) Estimated magnitudes versus USGS catalog magnitudes using the estimated locations after waveform amplitude correction by dividing the amplitude by a factor of 1.6 that found by best matching the USGS catalog magnitude. The mean of the residual is 0.1, and standard deviation is 0.4.

We could potentially correct this higher amplitude by scaling down the amplitude of the waveforms we recorded, estimating the best scaling factor to ensure that the MyShake data is as close as possible to USGS magnitudes. We did a grid-search with a factor from 1.0 to 3.0, and divided all the waveform amplitudes by using this factor and found that the factor 1.6 was the

best scaling factor. Figure 4.9b shows the estimation of the magnitudes with the scaled waveforms. We can see the results now change to mean 0.1 with a standard deviation 0.4 unit of the magnitude. Therefore, MyShake recordings are about 1.6 times larger on average in this dataset, and we could continue to calibrate the MyShake amplitude with more data recorded in the future.

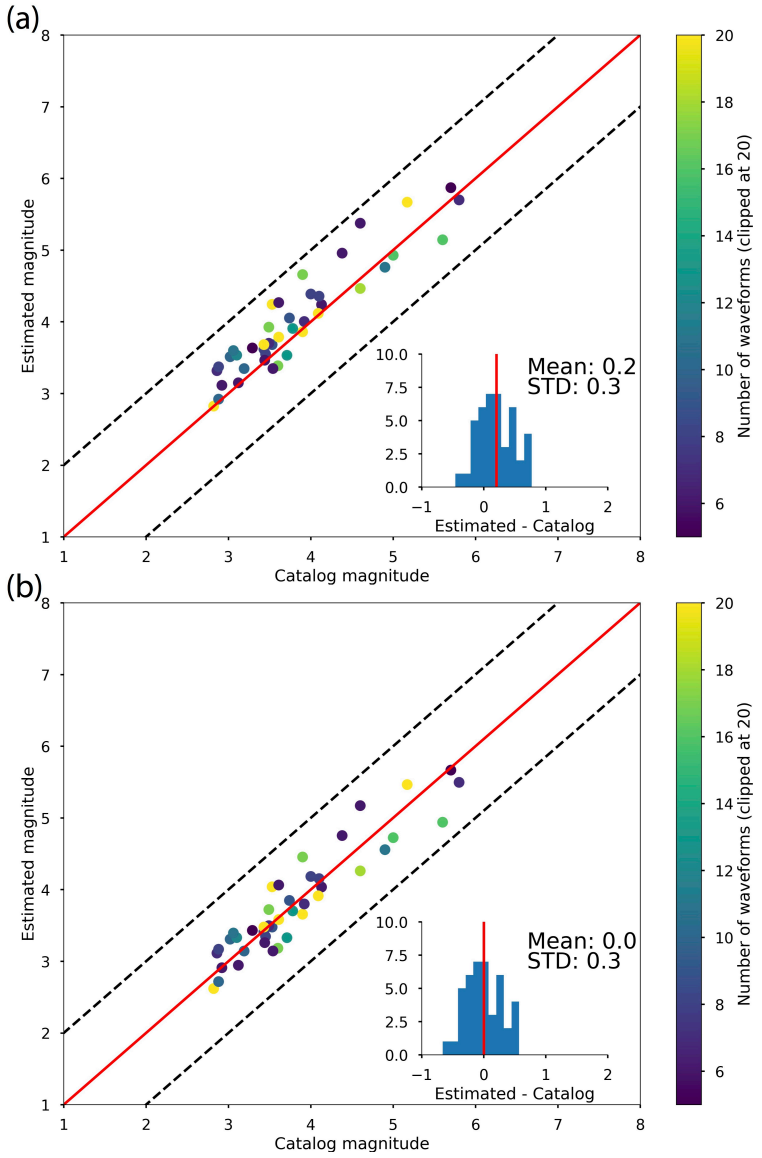


Figure 4.10: (a) Estimated magnitudes versus USGS catalog magnitudes using the USGS catalog locations. The red line is 1 to 1 magnitude, and the two dashed lines are 1 magnitude unit off from the 1 to 1 line. Each dot is estimated using different number of waveforms illustrated by the color. The inset plot is histogram of the residual magnitude (estimated magnitude – catalog magnitude). The mean of the residual is 0.2, and standard deviation is 0.3. The line in the histogram is the mean value. (b) Estimated magnitudes versus USGS catalog magnitudes using the USGS catalog locations after waveform amplitude correction by dividing the amplitude by a factor of 1.6 that found by best matching the USGS catalog magnitude. The mean of the residual is 0.0, and standard deviation is 0.3.

In order to see the difference of the magnitude estimation using the estimated locations and the USGS locations, we also estimated magnitudes using USGS catalog locations. Figure 4.10a shows estimated magnitudes using USGS catalog locations as the basis of calculation. It has a very similar distribution as that in Figure 4.9a using the estimated locations, with a slightly smaller mean of the residual as 0.2, and the standard deviation as 0.3. Figure 4.10b shows the estimation of the magnitude after the amplitude correction by dividing 1.6. The similar distribution illustrates that even using our own location estimation, with potential errors, we could still get a reasonable estimation of magnitude.

4.7 Discussion

The MyShake project is still an ongoing project, with many challenges that need to be addressed. However, the approach taken to build this global seismic network has been proved to be valid: within a short time, we have covered the whole globe, and the data collected shows great potential for routine seismological applications. There are still many things need to be explored to understand this type of network. In this discussion, we have described the difficulties, and the future improvements we think are required for creating a better global smartphone seismic network.

Firstly, on the operational side, many elements of this process currently still involve human interactions. For example, in our analysis, a seismologist needs to review the waveforms to confirm that they are useful, the S wave arrivals are selected manually, and so on. To build a fully functioning global seismic network, all these steps need to be automated. The smartphone seismic network differs from the traditional seismic network in many aspects. To name a few, the configuration of the smartphone network is constantly changing, and we have more detection capabilities at night than during the day (as more phones are steady during night); the triggering algorithm may be triggered on the P wave, S wave, or any part of the seismic motion; the instrumentation is highly heterogeneous in terms of the sensors inside the phones and different arrangements of the phone itself; and there is the ongoing question of how to increase the number of participants and keep the users using MyShake, etc. All of these complexities need careful treatments and new algorithms for automation. We are currently working on this issue and gaining better understanding of this new type of network.

Secondly, on the analysis side, there are a few limitations as regards using the MyShake data shown in this paper. As mentioned above, the qualities of the waveforms are different, even with two different phones at the same location; there is the question of how to weight the quality of the waveforms, and conducting quality control is not an easy problem. In addition, as regards events with many phones at further distances, we have to live with the fact that many of them only have the S wave arrivals available at the site, even though if we had P phases we could do a better job. Therefore, developing a robust phase selection algorithm for this noisy dataset is important. In addition, phones may be located in different types of buildings, on different floors, in places where the amplifications from the site response, from the response of the buildings, from the response of the desks, etc., will require us to use many phones in the region to aggregate the results in an average sense, and to calibrate the amplitude of MyShake recordings against the traditional seismic stations. Also, currently, the sources of location error estimates need more analysis, which can only be conducted once we have more detected events. Estimating location uncertainties from MyShake and quantifying each type of error will be useful for us in providing better location estimations.

Thirdly, for MyShake, due to the nature of the citizen science, there is a well-known limitation for building the seismic network. This seismic network is dynamically changing all the time, since the users are moving all the time. Additionally, MyShake users will tend to be clustered in large cities where populations are concentrated. Therefore, earthquakes in rural areas or far away from population centers can be missed. During the day, MyShake users are moving around, which can reduce the detection capability of the network. A side effect of this is that arthquakes occurring far away from cities may be detected via waveforms reported from several cities nearby, which can cause station coverage to be highly irregular, i.e. the earthquakes are outside of our seismic network. For this type of earthquake, usually the location estimations are not good, even on the traditional seismic network. A feasible solution to the above limitations is to combine information from the nearby traditional seismic stations or some permanent low cost sensors (Cochran et al. 2009; Clayton et al. 2011; Wu et al. 2016, Nof et al. 2017), which requires standardization of the data format and building a platform to conduct the data fusion.

4.8 Conclusion

In this paper, we show some basic understandings gleaned from the data we have collected through the MyShake network. We compared the magnitude distribution of MyShake detections with those from catalog events, which provide us with a good comparison of the capabilities of this smartphone seismic network. Additionally, from the waveforms we recorded from global users, we derived a relationship that could estimate for a certain magnitude, to what distance we expect to see recordings from the smartphone users. Comparing location errors and timing errors from our experiment and database helped us to evaluate the errors associated with these types of devices.

In addition, our analysis shows how to use the waveforms recorded by consumer smartphones to conduct routine seismological applications, including estimating the location, origin time and magnitude of an earthquake. From the estimated earthquake parameters, we could see the potential of the MyShake network to contribute to the current seismology community by providing more data to define these earthquakes. Especially in places where there are few traditional seismic stations, but a large population, MyShake could potentially provide valuable data to help understand earthquakes and tectonic settings. There are of course many challenges and limitations to address and overcome in the future, but a network such as MyShake can enhance our ability to better understand the earthquakes occurring globally, as well as to engage the public in locations where these earthquake occur.

4.9 Supplementary Figures

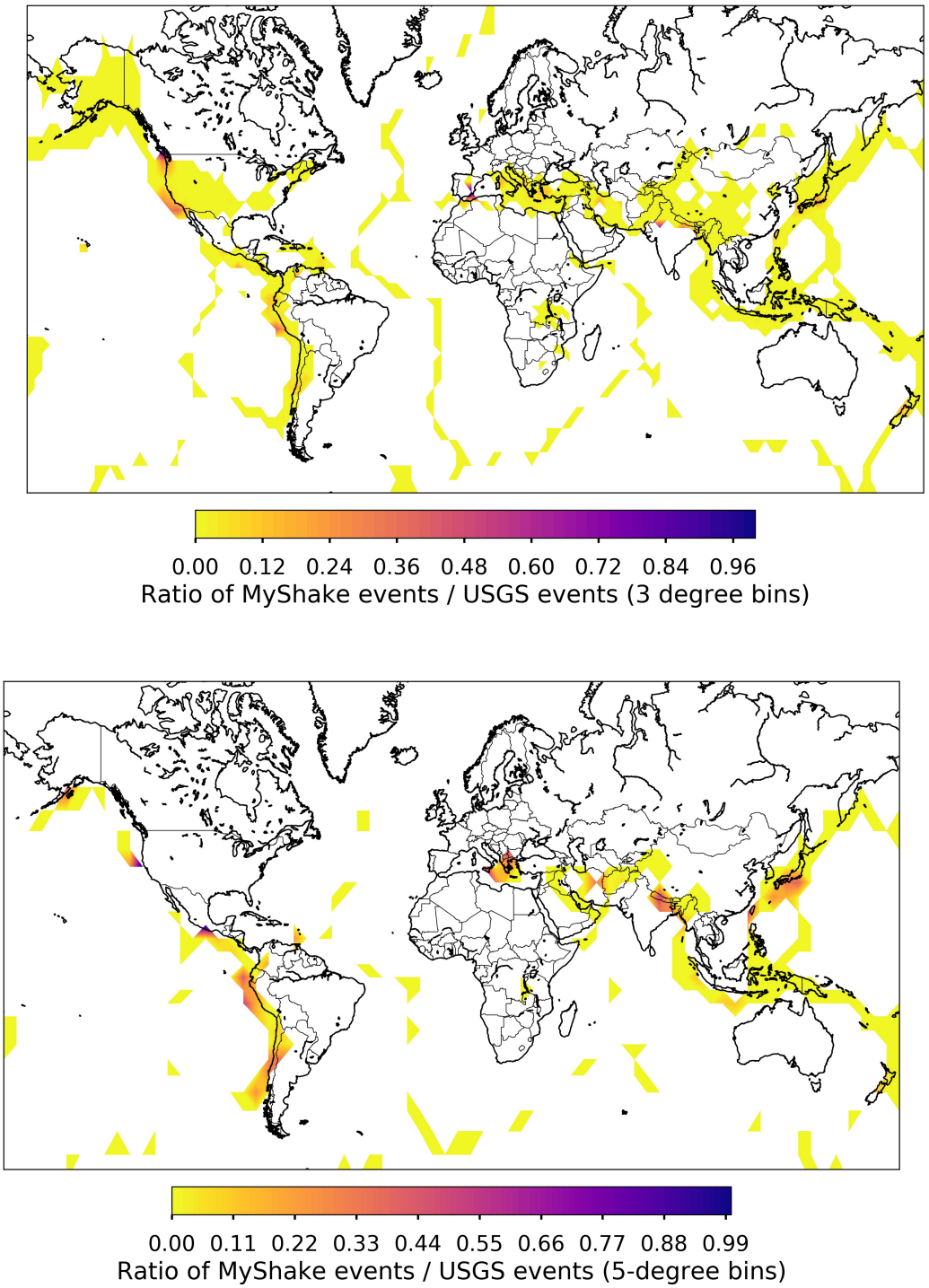


Figure 4.11: (a) The ratio of number of MyShake recorded events over that from USGS catalog within 3-degree bins. It includes all events above M2.5. The colors are showing the ratio. (b) Same as (a), but it shows all the events above M5.0 using a 5-degree bin.

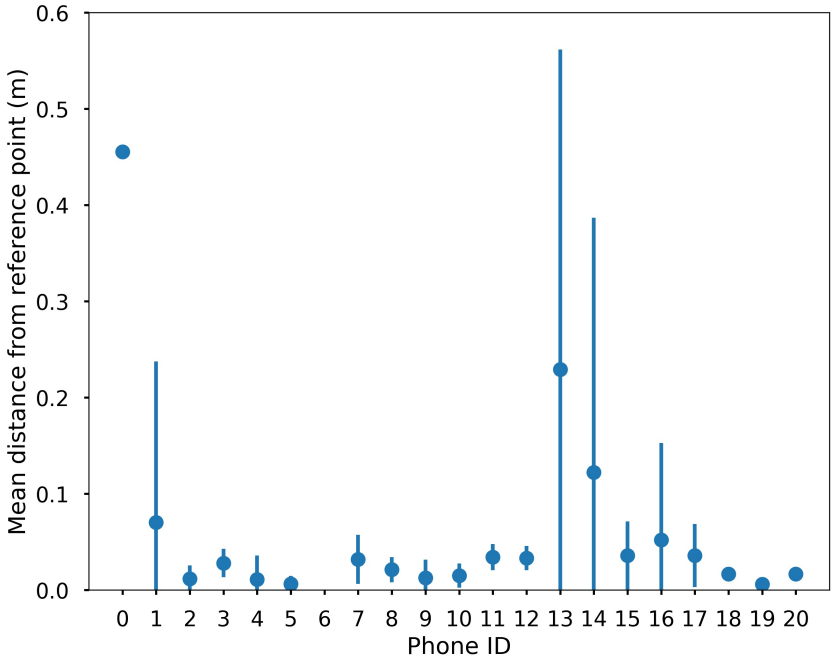


Figure 4.12: Location errors from the smartphone GPS experiments. The error bars are showing the standard deviations from all the measurements for each phone.

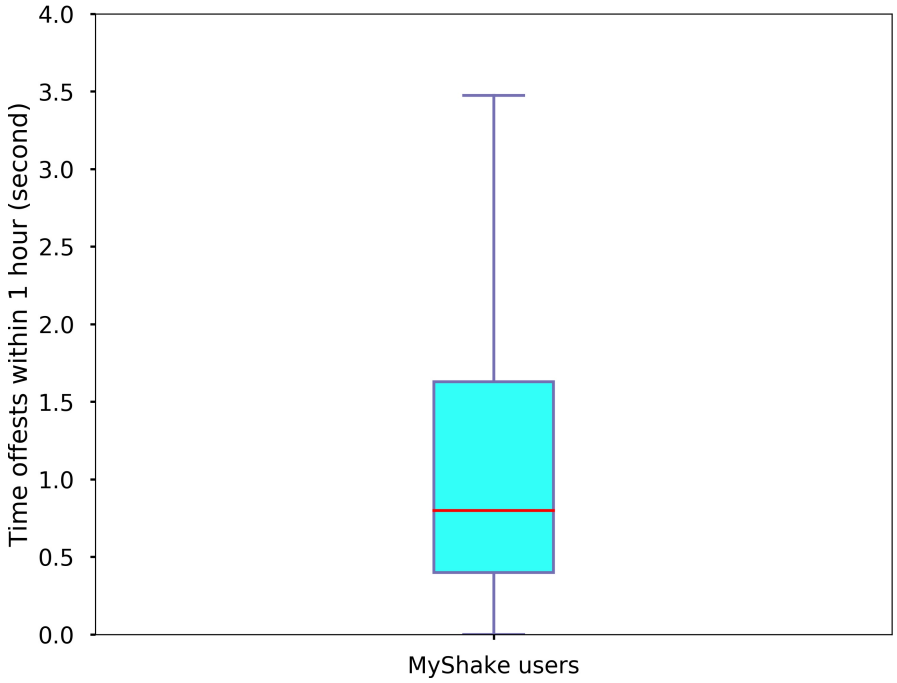
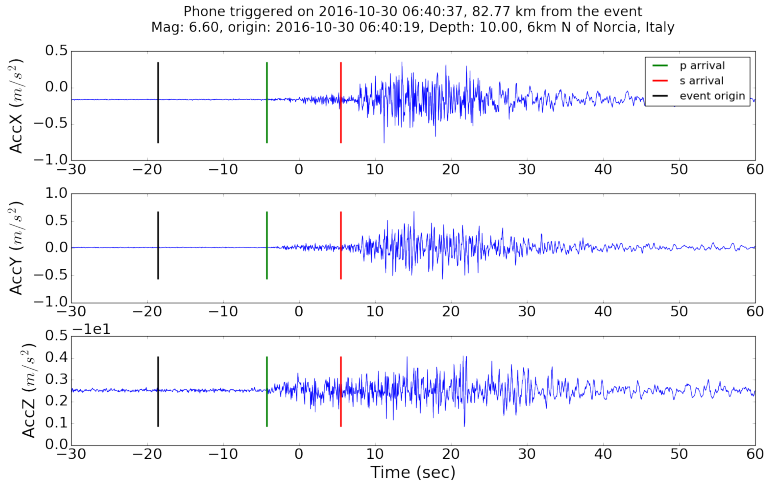
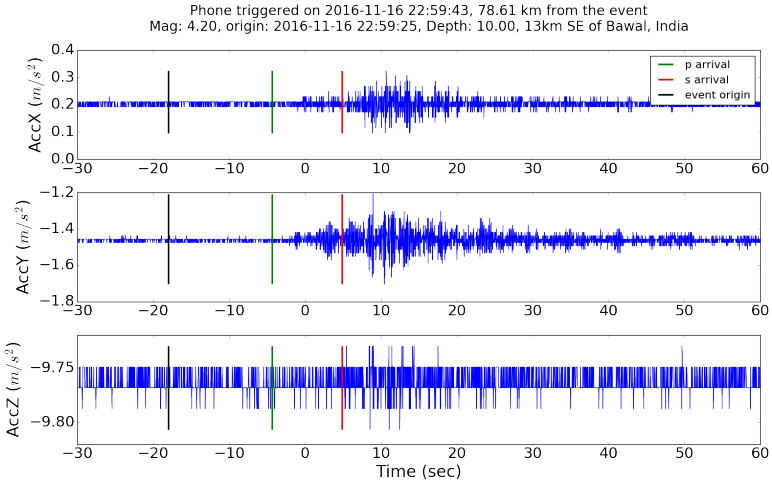


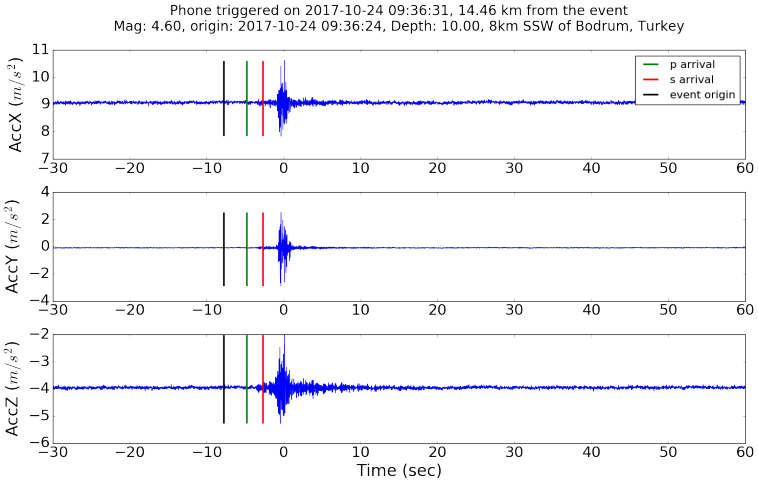
Figure 4.13: Timing error for the phones within 1 hour estimated by the offsets from the NPT synchronization. The red line is showing the median of the data and the box showing the 25 and

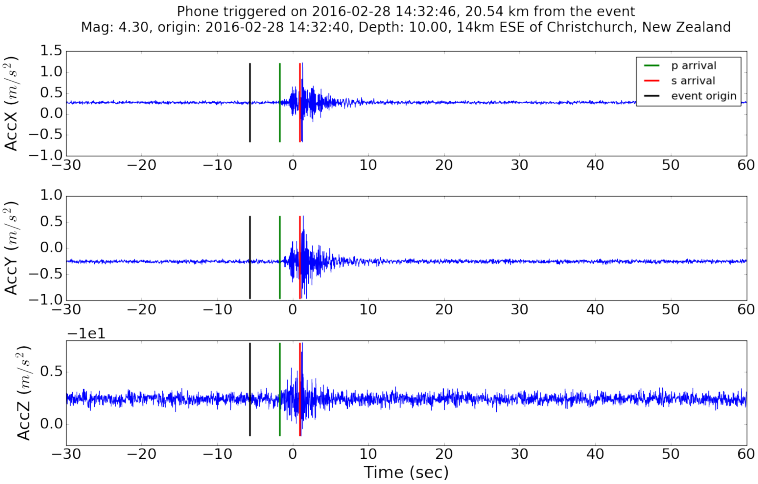
75 percentile, while the two bars at the top and bottom are whiskers which is 1.5 times of the Interquartile Range (IQR) from the box.

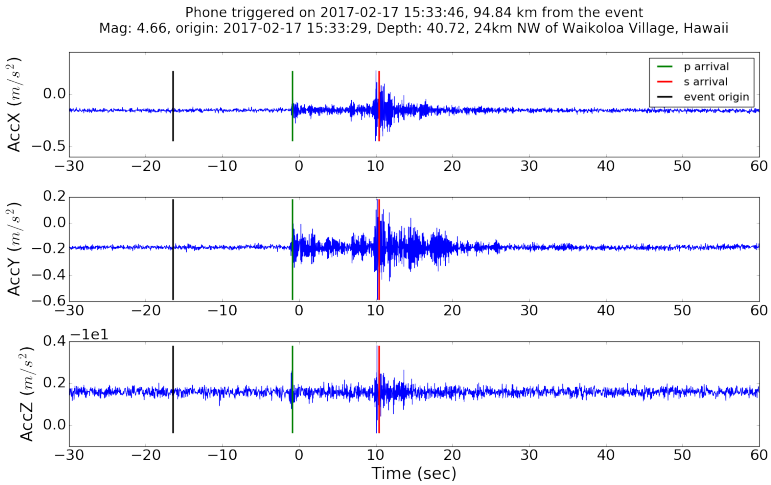
The following figures are selected 3-component acceleration waveforms from MyShake detections globally. The black line is the event origin time from USGS catalog, green and red lines are estimated P and S arrival time from model ak135. Time 0 on each panel is the time when the phone thinks it detects an earthquake.

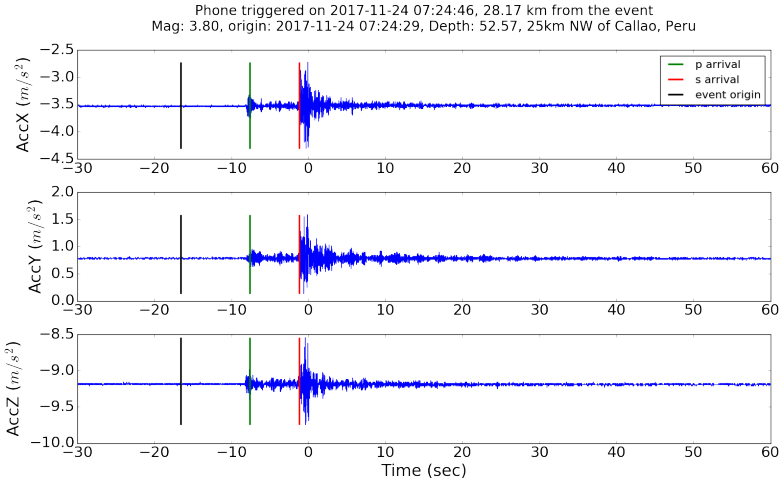


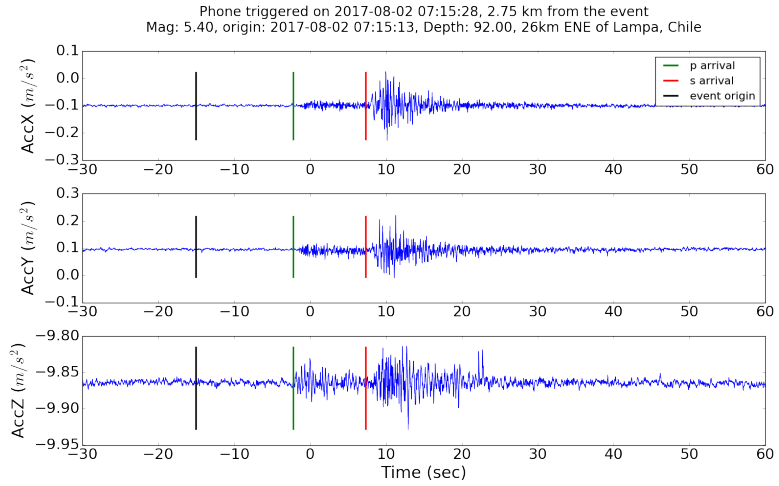


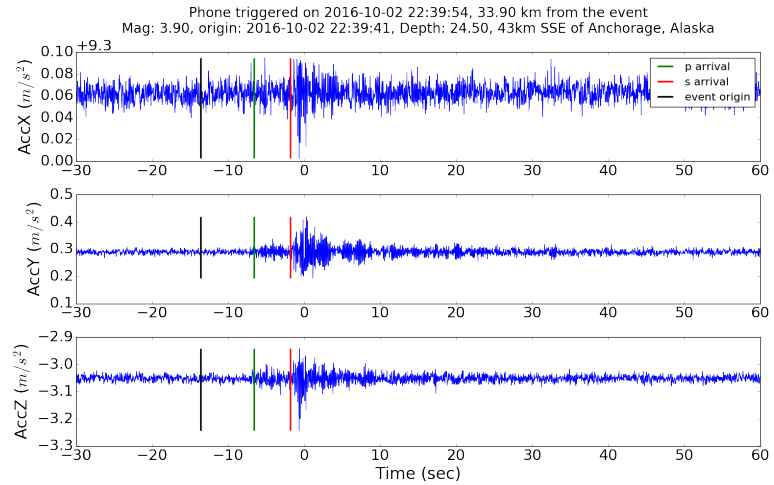

Phone triggered on 2017-10-24 09:36:31, 14.46 km from the event
Mag: 4.60, origin: 2017-10-24 09:36:24, Depth: 10.00, 8km SSW of Bodrum, Turkey
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)

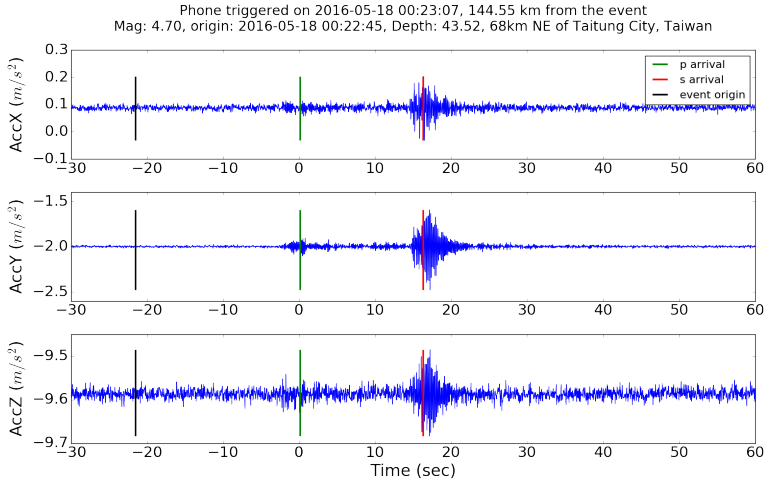

Phone triggered on 2016-02-28 14:32:46, 20.54 km from the event
Mag: 4.30, origin: 2016-02-28 14:32:40, Depth: 10.00, 14km ESE of Christchurch, New Zealand
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
−1e1
Time (sec)

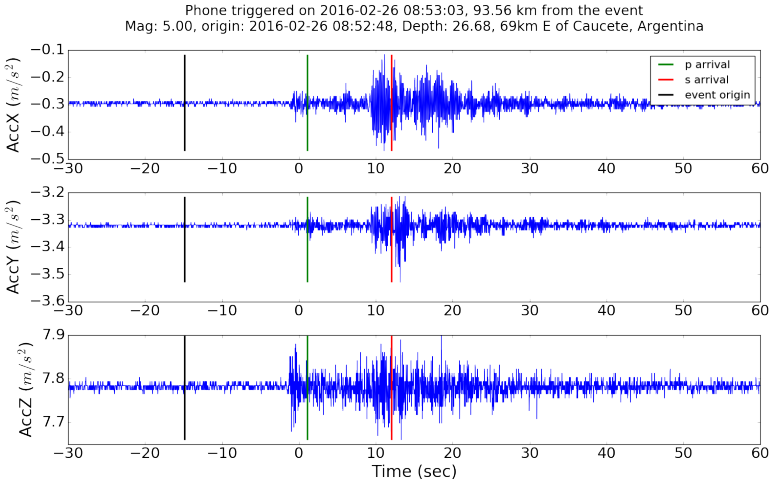

Phone triggered on 2017-02-17 15:33:46, 94.84 km from the event
Mag: 4.66, origin: 2017-02-17 15:33:29, Depth: 40.72, 24km NW of Waikoloa Village, Hawaii
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
−1e1
Time (sec)


Phone triggered on 2017-11-24 07:24:46, 28.17 km from the event
Mag: 3.80, origin: 2017-11-24 07:24:29, Depth: 52.57, 25km NW of Callao, Peru
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)


Phone triggered on 2017-08-02 07:15:28, 2.75 km from the event
Mag: 5.40, origin: 2017-08-02 07:15:13, Depth: 92.00, 26km ENE of Lampa, Chile
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)


Phone triggered on 2016-10-02 22:39:54, 33.90 km from the event
Mag: 3.90, origin: 2016-10-02 22:39:41, Depth: 24.50, 43km SSE of Anchorage, Alaska
+9.3
p arrival
s arrival
event origin
AccX (m/s^2)
AccY (m/s^2)
AccZ (m/s^2)
Time (sec)





4.10 Data and Resources

USGS catalog can be accessed at: https://earthquake.usgs.gov/fdsnws/event/1/. MyShake data are currently archived at Berkeley Seismology Laboratory and are constrained by the privacy policy of MyShake (see http://myshake.berkeley.edu/privacy-policy/index.html). Information about access to all the other data for research purposes contact rallen@berkeley.edu.

4.11 Acknowledgements

MyShake is a joint collaboration between the Berkeley Seismology Laboratory and Deutsche Telecom Silicone Valley Innovation Center. The Gordon and Betty Moore Foundation fund this analysis through grant GBMF5230 to UC Berkeley. We thank the MyShake team members: Roman Baumgaertner, Garner Lee, Arno Puder, Louis Schreier, Stephen Allen, Stephen Thompson, Jennifer Strauss, Kaylin Rochford, Doug Neuhauser, Stephane Zuzlewski, Sarina Patel and Jennifer Taggart for keeping this Project running and growing. All the analysis

of this project is done in Python, particularly the ObsPy package. The ANSS catalog data for this study were accessed through the Northern California Earthquake Data Center (NCEDC), doi:10.7932/NCEDC. We also thank all the MyShake users who contribute to the project!

Chapter 5 The potential of using smartphones for structural health monitoring

Published as: Qingkai Kong, Richard Allen, Julian Bunn, Monica Kohler, Thomas Heaton. "Structural health monitoring of buildings using crowdsourced smartphones", Seismological Research Letters, doi:10.1785/0220170111

5.1 Abstract

This chapter presents the results of a shaker test of the Millikan Library in Pasadena, California using sensors inside smartphones to demonstrate their potential usage as a way to monitor health states of buildings. This approach to structural health monitoring could allow many more commercial and residential buildings to be monitored because it removes the cost prohibitive nature of traditional seismic arrays, and the complexity of deploying the instruments. Recordings from the smartphones during the shaking show high correlation with those from a reference sensor in the building, illustrating that the phones can capture the shaking even when not fully coupled to the floor. The fundamental translational frequencies for the east-west and north-south directions and the torsional frequencies of the building can be extracted from single phone recordings. As we compare the displacement derived from the phone recording by double integration to that from the reference sensor, both phase and amplitude match well. Signal-to-noise is improved further by stacking records from multiple phones. These test results demonstrate the ability to extract the fundamental translational and torsional frequencies, and absolute displacements from upper levels of buildings shaken by small, local earthquakes. This work builds on the ongoing MyShake project – a global smartphone seismic network.

5.2 Introduction

Structures such as buildings, bridges, roads, and dams are an essential part of modern society. Even though they are designed to be used in different conditions, extreme unpredictable events, i.e., earthquakes, hurricanes, as well as deterioration due to aging, can cause serious concerns about the safety and functionality of the structures. Therefore, there is a need to monitor the health states of the structures throughout their lifetime.

Structural Health Monitoring (SHM) is a process that involves, first, observing a structural or mechanical system over time using periodically spaced measurements; second, extracting the damage-sensitive features from these measurements; and, third, statistically analyzing these features to determine the current state of system health [Farrar and Worden, 2007]. Different types of sensors are deployed in the structures either permanently or temporarily to extract measurements such as acceleration, velocity, displacement, deformation, stress, temperature etc. [Moreno-Gomez et al., 2017]. SHM is expensive both in terms of the hardware and the human effort; it is implemented in only a few large-scale structures, and must currently be deployed and maintained by practicing structural engineering professionals. Therefore, it is virtually impossible to conduct continuous, long-term monitoring of the state of health of most buildings due to the cost of both hardware and human efforts.

The emergence of wireless sensors, low-cost MEMS (Micro Electro Mechanical Systems) sensors and sensor networks has started to provide lower cost solutions to replace traditional tethered monitoring systems [Xu et al., 2004; Paek et al., 2005; Lynch and Loh, 2006; Kim et al., 2007; Kohler et al., 2013; Yin et al., 2016]. But to use them in a nationwide effort or even at the city level would require a large-scale effort for engineers to deploy and maintain these systems. In addition, many building owners are reluctant to install sensors in their buildings as they are fearful of legal issues or simply choose not to prioritize their building performance problems.

MyShake aims to build a global smartphone seismic network by utilizing the power of crowdsourcing. It turns an everyday hand-held device into a portable seismometer by monitoring data from the accelerometer in a smartphone to detect earthquakes [Kong et al., 2015, 2016a]. After release of the MyShake app to the public on Feb 12th 2016, it has been downloaded by more than 270,000 users globally. Today, about 10,000 active phones contribute data to the system each day for monitoring earthquakes. The results from the collected data are promising [Kong et al., 2016b]; MyShake can record M5 earthquakes up to about 200 km from the smartphone, and it can record small magnitude earthquakes (M2.5) at closer distances.

The quality of the data recorded by MyShake, and the ease of building and scaling up with this network, led us to investigate whether we might expand the use of MyShake data to the area of SHM. If private smartphone sensors can be used as a mechanism to collect data on the health state of buildings, then smartphones could overcome the substantial challenge of deploying the sensors manually in buildings, and provide a way to monitor the buildings at very low cost of hardware and maintenance. Since the smartphones may also be located throughout a building, this approach could provide a dense in-building network to monitor the structural health state of the building floor by floor.

Multiple groups have studied the feasibility of using a low-cost sensor network to conduct SHM [*Cochran et al.*, 2009b; *Clayton et al.*, 2011, 2015, *Kohler et al.*, 2013, 2016; *Yin et al.*, 2016]. These studies use specially designed low-cost sensor boxes that can be deployed in a building either by professional engineers or community volunteers. These sensors cost tens of dollars to a few hundred dollars each, but the more significant cost is associated with the fact that someone must manually deploy the sensors in buildings. This makes it challenging to scale up these networks due to the human labor, permitting, and permission efforts required.

In our study, smartphones provide an opportunity to replace this manual deployment process with a straightforward software download and installation onto the user's phone. In a typical smartphone, a built-in 3-axis accelerometer measures the movement of the phone in 3 dimensions (one vertical and two orthogonal horizontal components if the phone is oriented with one side parallel to the ground). Previous shake table tests of the accelerometers inside smartphones have shown that typical high-end accelerometers (used in iPhones and high-end Android phones) are capable of recording motion in the frequency range 0.2 – 20 Hz and amplitude range 10 – 2000 mg (g is gravitational acceleration) [D'Alessandro and D'Anna, 2013; Reilly et al., 2013; Dashti et al., 2014; Kong et al., 2016a]. Engineers have explored the use of smartphone sensors to monitor the health states of large-scale structures such as bridges. Yu et al., 2015 conducted a series of tests to show that using smartphones to carry out health monitoring of bridges is feasible. [Ozer et al., 2015; Feng et al., 2016] show that by using

smartphones attached to a pedestrian bridge, they can infer modal parameters from the recordings of the phones and layout of a structure for a crowdsourcing platform.

Here we explore the use of private smartphones to monitor building health. This is the first step in investigating whether the existing global MyShake network of smartphones as well as similar crowd-sourcing smartphone efforts [Faulkner et al., 2014; Finazzi, 2016] could be harnessed for this purpose, in addition to monitoring earthquake activity. We show results from a shaker test of the Millikan Library building on the Caltech campus in which we placed the phones on the floor of the 9th (top level) floor. We determine if we can extract the fundamental frequencies of the building, and estimate the displacement of the floor due to motions similar to that from a small, nearby earthquake. Personal smartphones are of course in motion with their owners for portions of the day as the owner walks, commutes etc. Our test is similar to the results expected for stationary phones resting on stands at night, or when placed on a desk or left in a bag on the floor. The results presented here illustrate the potential of using MyShake-enabled personal smartphones to record building shaking resulting from nearby earthquakes and using that data to extract the building characteristics. We also present a method to determine the orientation of the smartphone if its orientation is not known, but prior information about the building characteristics is available.

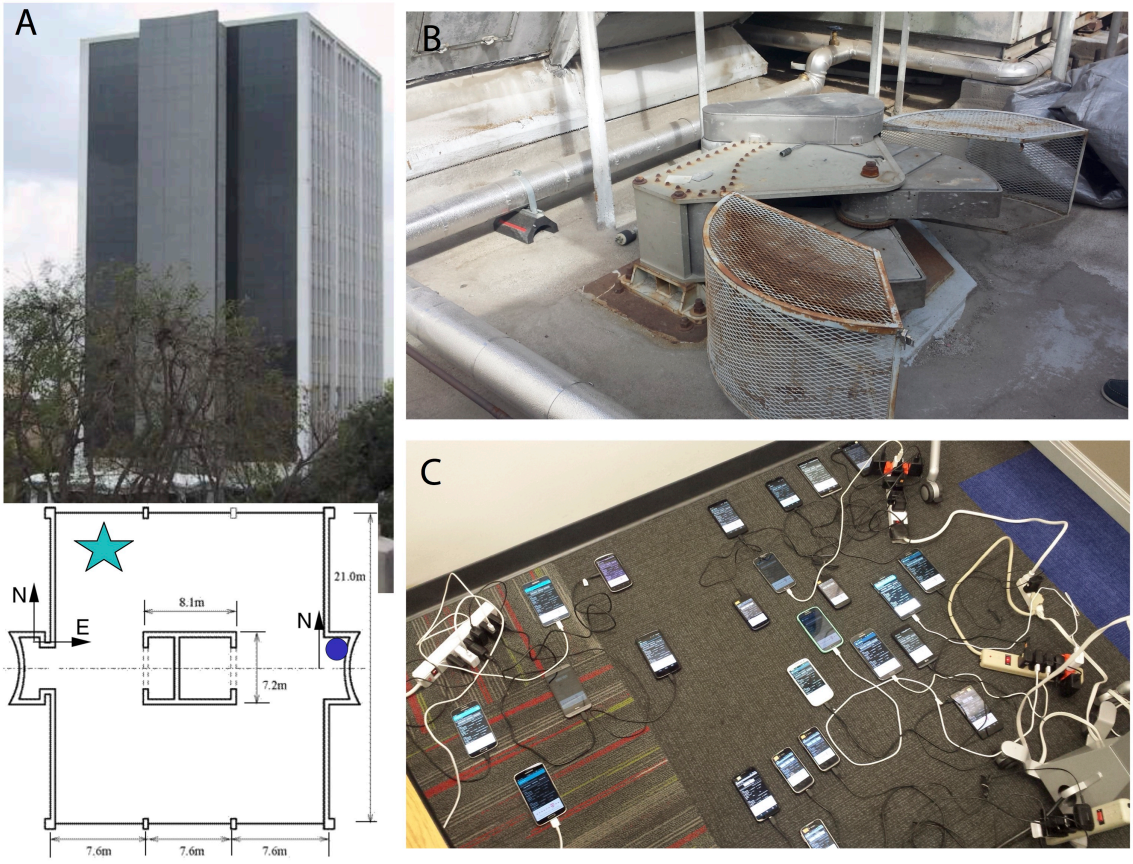


Figure 5.1: Background of the test. (A) The Millikan Library building viewed from the northeast. The two dark colored panels on the near-side of the building comprise the east shear wall

(modified from [Bradford et al., 2004]). The inset figure (modified from [Clinton et al., 2006]) shows the plan of the building, where the star is the location of the test phones and the dot is the location of the Episensor station CI MIK. (B) The shaker located on the roof, used to generate oscillation of the building. The two exposed buckets contain lead masses which spin in opposite directions to generate a sinusoidal horizontal force. (C) The 25 smartphones used in the test, all placed on the floor of the 9th (top level) floor. The duration of the shaking in the North-South direction is 1:29 pm to 3:02 pm, and in the East-West direction is 3:38 pm to 5:03 pm, local time.

5.3 Background of Millikan Library

The shaker test was conducted at the Millikan Library building on the campus of the California Institute of Technology (Caltech) (Figure 5.1a). Millikan Library is a nine-story, reinforced concrete building, approximately 44 m tall, and 21 m by 23 m in plan. The building has concrete moment frames in both the east-west (E-W) and north-south (N-S) directions. Shear walls on the east and west sides of the building provide most of the stiffness in the N-S direction, and shear walls in the central core provide added stiffness in both directions [Bradford et al., 2004].

The Millikan Library building is instrumented with a permanent, dense array of uni-axial strong-motion sensors (Kinemetrics FBA-11s in 1 g and 2 g) with 36 channels throughout the building. On each floor there are 3 horizontal accelerometers; in addition, 3 vertical accelerometers are installed in the basement. A 3-axis Episensor is installed on the 9th floor with a 24-bit data logger [Bradford et al., 2004] (see the location in Figure 5.1a). In addition, the building is instrumented with 10 Community Seismic Network (CSN) accelerometers distributed on per floor [Kohler et al., 2013]. Previous studies have shown changes in the modal parameters of the building through time as a result of large-amplitude ground shaking [Clinton et al., 2006]. These studies illustrate the importance of understanding and documenting the dynamic properties of different classes of structures (i.e. steel-frame versus reinforced-concrete buildings) within the linear response regime before nonlinear response might occur.

5.4 Method

A Kinemetrics model VG-1 synchronized vibration generator (”shaker”) was installed on the roof of Millikan Library in 1972 (Figure 5.1b). The shaker has two buckets that rotate elliptically in opposite directions around a center spindle. These buckets can be loaded with different configurations of lead weights, and depending on the alignment of the buckets, the shaker can apply a sinusoidal force in any horizontal direction [Bradford et al., 2004].

In our tests, we applied forces to the building first in the N-S direction, and then in the E-W direction at discrete frequencies. The applied frequencies of oscillation spanned 0.2 to 2.45 Hz and were varied gradually over the course of approximately 2.5 hours. The rate of change of frequency varied over the course of the sweep. The sweep progressed at 0.05-Hz intervals with frequencies held constant for 60 s most of the time, but near the modal frequencies, the constant-frequency interval was extended to 600 s. The extended run-time was used for the building’s fundamental east-west mode frequency of 1.2 Hz, the fundamental north-south mode frequency of 1.7 Hz, and the fundamental torsional mode frequency of 2.4 Hz. Additional details of the test runs are provided in the supplementary material (Table 5.1).

Brand/Model	N	Acc. Type	Vendor	Max Range (m/s^2)	Resolution (m/s^2)
Samsung S3	1	MPL	InvenSense	19.6	3.83E-02
Samsung S4*	2	K330	STMicroelectronics	19.6	5.99E-04
Samsung S5*	1	MPU6500	InvenSense	19.6	5.99E-04
Sumsung S6*	1	MPU6500	InvenSense	19.6	5.99E-04
Sumsung Note2	2	LSM330DLC	STMicroelectronics	19.6	9.58E-03
Samsung Note4	1	LCM20610	InvenSense	39.2	1.20E-03
Samsung Note5*	1	K6DS3TR	STMicroelectronics	39.2	1.20E-03
Samsung-Exhibit-II	1	BMA222	Bosch	19.6	1.53E-01
Nexus 5	2	MPU6515	InvenSense	39.2	1.20E-03
LG-G2*	1	LGE	STMicroelectronics	39.2	1.20E-03
LG-G-Stylo	1	LGE	Bosch	156.9	9.58E-03
LG-Leon*	1	LGE	Bosch	156.9	9.58E-03
LG-G4	1	LGE	Bosch	156.9	9.58E-03
HTC-One-M9	1	N/A	HTC Corp	39.2	1.00E-02
MotoX	2	LIS3DH	STMicroelectronics	156.9	4.79E-03
HuaWei Prism	3	BMA150	Bosch	39.2	1.53E-01
Sony Xperia	2	MPL	InvenSense	19.6	3.83E-02
HTC Amaze	1	Panasonic	Panasonic	19.6	1.20E-02

Table 5.1: Smartphones used in the test. Brand and model of the phones are shown, N is the number phones used in the test. Acc. Type is the model of the accelerometer. Max Range and Resolution show the range of the amplitude and the smallest measurable value that the sensor can measure. The phones flagged with * indicate that the phone is used in the 7-phone stack. The resolution values are from the sensor specifications.

Smartphones were placed in the northwest corner of the 9th floor (top level) of Millikan Library (Figure 5.1a, c) with their x-axis approximately aligned in the E-W direction. Note that for this test the phones are placed at the most advantageous location, as the displacements on the corner of the 9th floor will be greater than at many other locations in the structure. Twenty-five different models of Android phones were tested (see Table 5.1 for details) as the accelerometers in Android phones are of various qualities [Kong et al., 2016a]. The phones have flat response in the frequency range 0.1 to 12.5 Hz, and we use a sampling rate of 25 samples per second for these tests. The resolution listed in Table 1 is from the phone specifications and shows the best case when the phone is insulated from environmental vibrations.

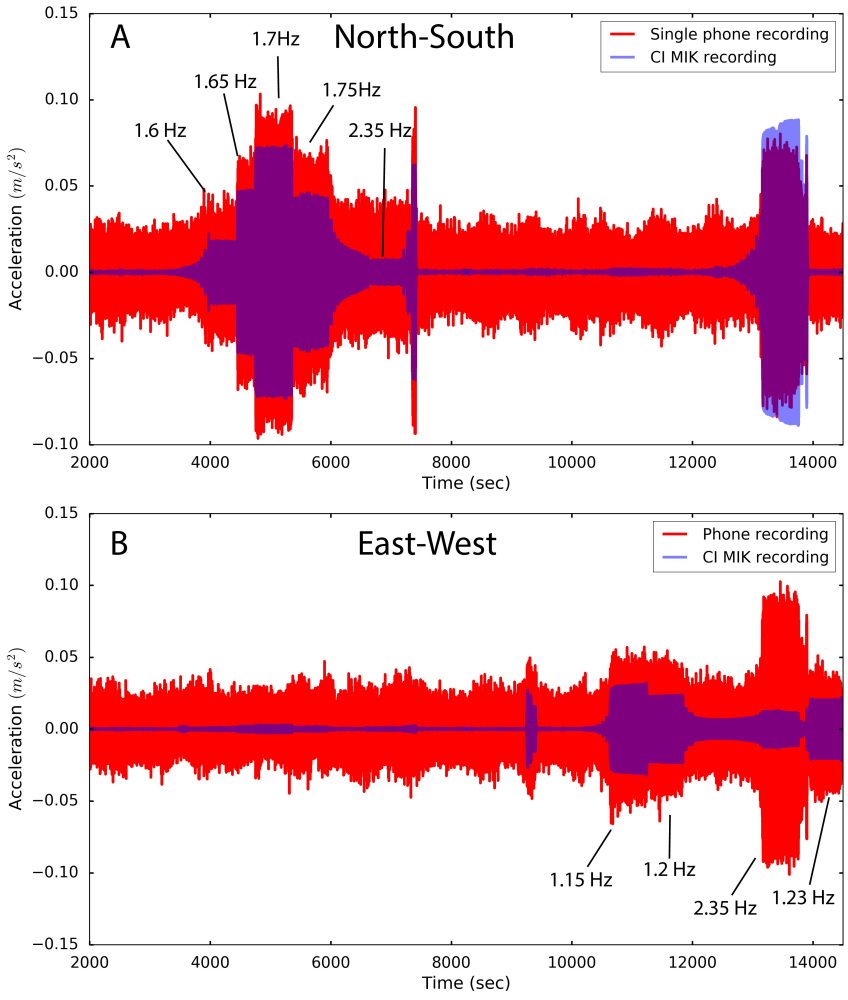


Figure 5.2: Waveform comparisons. Waveform comparisons between the Episensor (CI MIK) and a Samsung Galaxy S4 phone for the two horizontal components. (A) North-south component. (B) East-west component. The red time series is from the single phone recording (Samsung Galaxy S4), and the blue is from the MIK recording. Frequency labels indicate when the test run is at or near the fundamental or torsional frequencies of the building. The amplitude and phase alignment is generally good; see **Figure 5.8** which expands the time window of 4820 to 4850s.

5.5 Results

Acceleration waveforms were recorded by the smartphones and compared to waveforms from a 3-component Episensor which is permanently installed on the 9th floor, to validate the smartphone motion. The Episensor has a 24-bit Q980 data logger, and its data are continuously telemetered to the Southern California Seismic Network (SCSN) as station MIK. We will refer to it as station CI MIK in the following figures. Comparison of the acceleration waveforms between the Episensor and one of the Samsung Galaxy S4 phones is shown in Figure 5.2. The recording from the phone generally has a larger amplitude during peak shaking and a higher noise level which is above the lower amplitude shaking compared with the recording from CI MIK. Between times 2000 to 8000 s the shaking was applied in the N-S direction, and from 9000 to 15,000 s it was applied in the E-W direction (Figure 5.2). During each time interval, shaking was applied at a range of frequencies including the fundamental translational mode in that direction and the fundamental torsional mode which excites motion in both directions due to the rotational nature of torsion about a vertical axis. It is clear that the recordings from the phone and the Episensor from N-S and E-W shaking show good correlation, though the amplitude of the phone signal is greater than the Episensor. A shorter time window of the comparison is shown in Figure 5.8 illustrating the phase matching. The signals during the torsional motion (shaking frequency around 2.35 Hz), however, do not correlate in the same way, due to the different locations of the sensors on the 9th floor. At the location of the Episensor, the torsional motion is mainly in the N-S direction; the smartphones located at the northwest corner of the building, however, experienced motion in both the N-S and E-W directions, causing the amplitude difference. This can be seen in the time range from 13000 s to 14000 s (Figure 5.2).

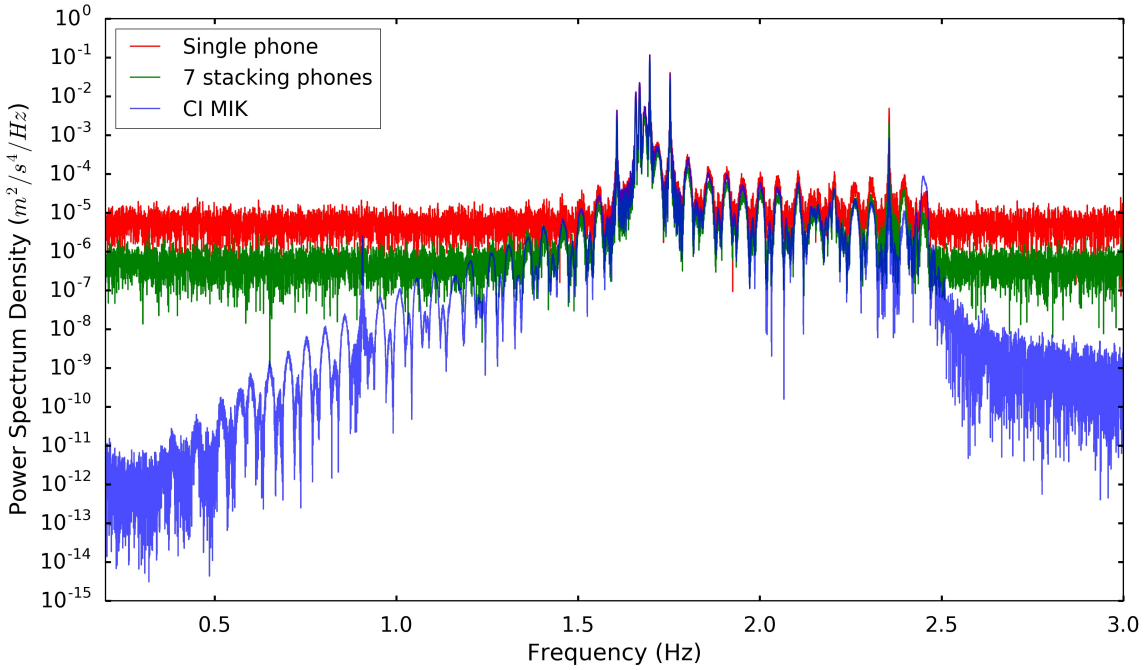


Figure 5.3: Spectrum comparisons in the N-S direction where the fundamental mode of the building is at 1.7 Hz and the fundamental torsion mode is at 2.35 Hz. Blue is spectrum for the

Episensor (CI MIK), red is spectrum for a single phone (Samsung Galaxy S4) and the green spectrum is from the 7-phone stack. The modal peaks are clearly visible in all cases.

Just one smartphone records building response to the shaking well, especially at the fundamental frequencies and torsional frequency. From the spectrum (Figure 5.3), we clearly see the peak at the fundamental frequency and torsional frequency, and can thus extract them from a single phone. Since we have 25 smartphones at the same location we can also stack them to improve the signal-to-noise ratio. This assumes that noise recorded by different phones is truly random, and stacking across different phones will cancel the noise but not the coherent signals caused by the building's response to shaking.

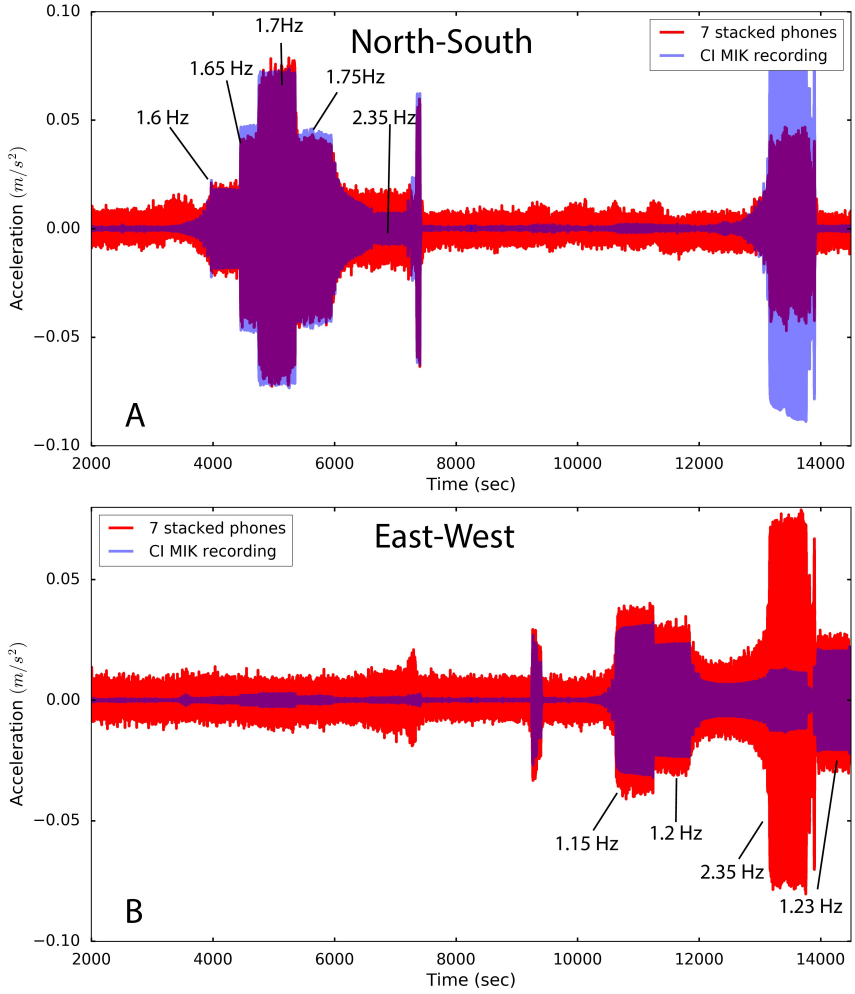


Figure 5.4: Waveform comparisons for two horizontal components between the stack of 7 phone recordings and the Episensor (CI MIK). (A) N-S component. (B) E-W component. The red time series is from the stacked recordings from 7 phones, and the blue is from the MIK recording. Frequency labels indicate when the test run is at or near the fundamental or torsional frequencies of the building.

Since the phones are not synchronized with each other (each phone has its own network time), we use cross-correlation to find the best alignment of the recordings from different phones. We calculated the cross-correlation of the entire N-S component time series between different phone recordings and a base phone recording by shifting them within 120 s windows to find the maximum correlation coefficient. In order to get the best stack, we select the phones with different thresholds of the correlation coefficients. Correlation coefficient thresholds of 0.6, 0.5, 0.4, 0.3, resulted in stacks with 7, 10, 14, and 16 phones, respectively. The quality of the waveforms recorded are also variable due to the different accelerometers in the phones; thus, stacking more phones does not always improve the resulting waveform. By comparing the waveforms and spectra (more on this below), we conclude that stacking 7 phones gives us the best results (the correlation coefficient is larger than 0.6). These 7 phones are flagged with a star in Table 1, and we see that most of them have the best accelerometers based on specification resolution. Figure 5.4 shows the comparison of the waveforms between the Episensor (CI MIK) and the stack of 7 phones. We see that the stack amplitudes are a better match to those observed on CI MIK. The noise levels during low-amplitude shaking are lower, and the observed amplitudes during peak shaking are more similar.

The main goal of our shaker test is to extract the fundamental frequencies of the building from the phones and to compare them for accuracy with those of the Episensor. To calculate the spectrum of the building's shaking, we found that a multi-taper spectrum analysis obtains better results than a direct FFT. In the multi-taper analysis, the time series data to be analyzed is multiplied by a series of orthogonal tapers, and then Fourier transformed and squared to obtain the estimate of the power spectrum density (PSD). The orthogonal tapers will generate many independent estimates of the PSD instead of only one, and an average of them will suppress the random variance in the estimation [Prieto et al., 2009]. We select the N-S and E-W component individually, and apply the multi-taper analysis. Figure 5.3 shows the amplitude spectrum for the recordings of the N-S component from the Episensor, a single phone, and the stacked phone time series. Overall, we observe that the single phone spectrum has a noise level around 10e-5 and the stacked phone time series has a noise level of about 10e-6, which make the peaks around 1.25 to 1.5 Hz observable. However, the fundamental frequency in the N-S direction is clearly visible in all cases at 1.7 Hz. The fundamental torsional mode frequency at 2.35 Hz is also distinguishable in all cases. The E-W component spectrum produces similar results and the peak of the fundamental frequency is clearly visible; see Figure 5.9 in the supplementary material for the spectrum comparison for the E-W component.

Peak and relative displacement amplitudes at a given floor of the building are also important to the civil engineering community to quantify localized deformation of the building (e.g., inter-story drift). In Figure 5.5, we show comparisons of the displacement time series between the Episensor, a single phone, and stacked phone recordings. Overall, the result from 7 stacked phone time series shows better agreement with the Episensor in both phase and amplitude. These are obtained through double integration of the acceleration recordings by first removing the mean and trend in the record, and then applying a 0.5 Hz high-pass filter. They show good agreement with a peak displacement of about 0.05 cm for this frequency range.

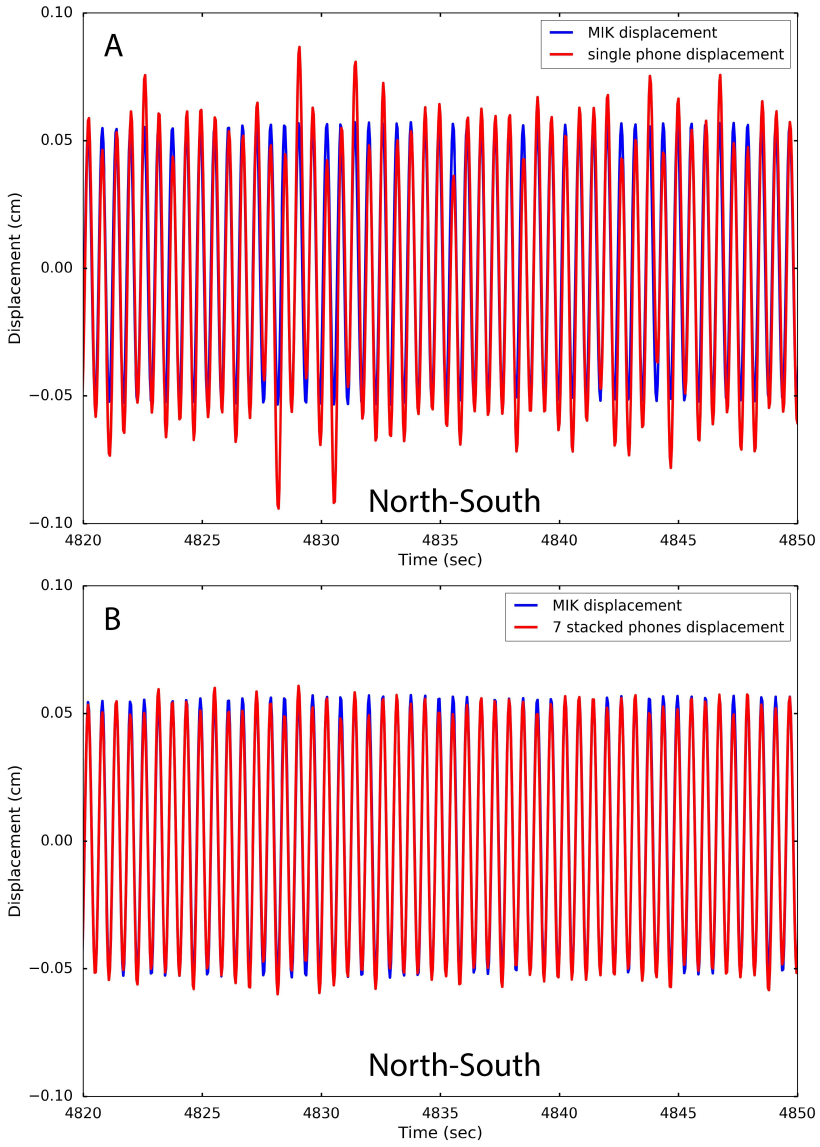


Figure 5.5: Displacement time series comparisons (high-pass filtered at 0.5 Hz). These recordings are extracted from the period of largest-amplitude response to the shaking that occurred in the N-S direction, between 4820 and 4850 s (See **Figure 5.2**a for reference). (A) The Episensor (CI MIK) compared with a single phone (Samsung Galaxy S4). (B) The Episensor compared with the 7-phone stack. See **Figure 5.8** for the accelerations in the same time range.

5.6 Estimation of the orientation of the phones

Sometimes we have the reverse problem in which we know the building's fundamental frequency but do not know the phone's orientation. If a building's modal frequencies and corresponding mode shapes are already known, then it is possible to deduce the orientation of an arbitrarily-rotated smart phone that is recording a known mode: rotate the phone until the resulting filtered motion is aligned with the mode shape. Analogous methods have been applied

to find the orientation of three-component seismic sensors in various conditions [Duennebier et al., 1987; Ekstrom and Busby, 2008; Stachnik et al., 2012]. In our case, application of this method will only work for the translational model shapes, and when the phone produces a high signal-to-noise ratio recording, most likely from the free oscillations of the building following earthquake shaking. Also, the elevation of the phone is a factor as phones on the higher floors will likely have higher signal-to-noise ratio recordings.

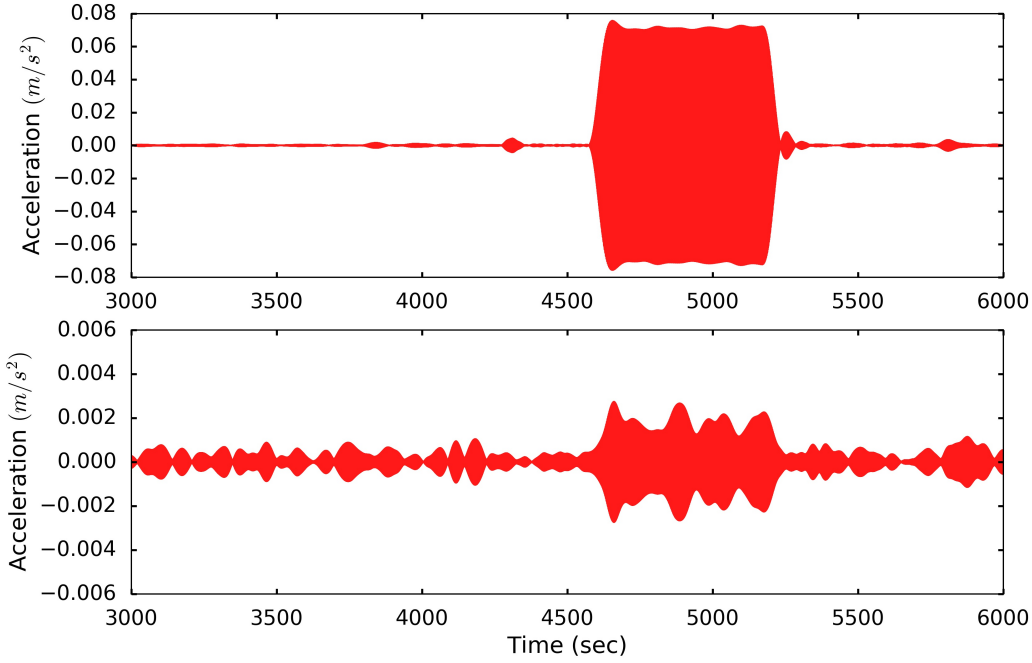


Figure 5.6: Horizontal component recordings after application of a narrow-band filter 1.69 to 1.71 Hz. Energy can be seen on both components. The upper panel is the component oriented roughly N-S, and bottom panel is the component oriented roughly E-W. The phone used here is the same Samsung Galaxy S4 as the one shown in **Figure 5.2**.

Since we know that the N-S fundamental frequency of the Millikan Library building is 1.7 Hz, we should be able to estimate which component is N-S by applying a narrowband filter to the phone's two horizontal component records. Figure 5.6 shows that the component shown in the top panel is likely close to the N-S direction; this is before any corrective rotation is applied. We observe that there is a small signal on the bottom panel component as well, likely due to the fact that the phone is not perfectly oriented N-S during the tests.

To find the most accurate orientation of the phone during the tests, we can rotate the two horizontal components until an angle is found that minimizes the signal on one component. Figure 5.7 shows the result of rotating the phone 1.5 degrees; the energy on the E-W component is minimized. Of course, this is easy for our test, since we shake the building in one direction at a time. During an earthquake, the later parts of the motion that are dominated by free vibrations should be usable in the same way to orient the phones. This will allow more meaningful analysis of earlier parts of the record.

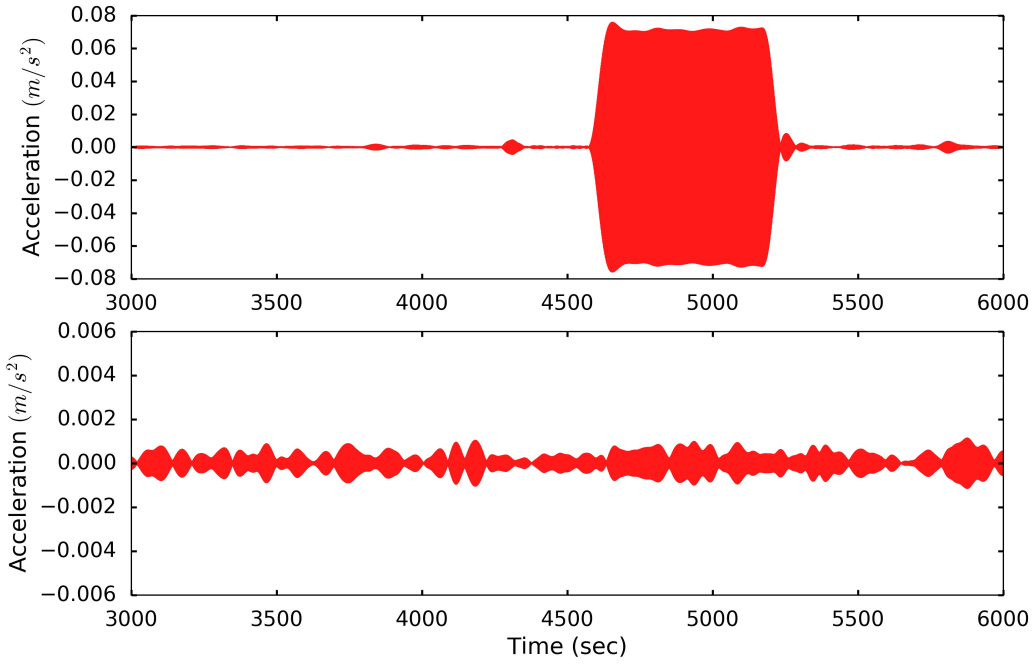


Figure 5.7: Result of rotating the horizontal components of the narrow-band filtered signal (**Figure 5.6**) anticlockwise 1.5 degrees in order to minimize energy on one component. The upper panel shows the axis aligned to N-S after having rotated the components, the lower panel is the axis aligned to E-W.

Although a building's mode may not already be known, in some cases it is possible to estimate modal properties from the known geometry of the building. In particular, many buildings are approximately rectangular and the orientation of the building's normal modes are approximately aligned with the natural axes of the building. If structural design information is available to determine the most compliant direction of building deformation, then one can assume that the lowest-frequency normal mode of a building is also aligned with this direction. Even if the axis of the lowest frequency mode is unknown, it is often safe to assume that it is one of the two natural axes of the building. Knowing which of the two axes is the correct one can be determined by observing the building's response with just one record of approximately-known orientation (it is most beneficial to have at least one sensor with known orientation).

In some cases, tall buildings may have additional modes (overtones) at frequencies that are odd-integer multiples of the fundamental mode frequency [Lee et al., 2003]. The orientations of higher modes can also be used to orient the phones. Finally, there are cases in which shear waves travel vertically up a building (typical wave velocity of 150 m/s). In general, the polarization of vertically propagating shear waves is approximately constant over the building height. If the polarization of the incident wave is known, then the orientation of the phones can be determined by finding the appropriate rotation to reproduce the incident ground motion. Examples of how this is applied can be found in [Cheng et al., 2015].

5.7 Discussion and Conclusions

Structural health monitoring is important for keeping track of the changes in the state of

buildings, not only of large-scale structures, but also of everyday residential buildings for the purpose of damage detection. Harnessing the smartphones of individual private users could open the door to monitoring many more structures including residential buildings in the future.

When compared with the current wireless sensor network monitoring systems, using consumer smartphones as a way to conduct structural health monitoring has the following benefits:

- Ability to monitor millions of buildings within short periods of time.
- Almost no cost for hardware and installation labor.
- Low cost for long-term maintenance.
- Complementary data to the current monitoring system.

The first point is the most important benefit of using smartphones; millions of buildings could be monitored with just the download of an application. This will greatly improve our current monitoring ability at the city scale or even nation scale to reduce the earthquake risks. Also, a smartphone monitoring system is complementary to existing structural health monitoring systems by providing more data in the same building for validation and to fill in spatial sampling gaps.

This chapter is only a starting point for the concept of crowdsourcing structural health monitoring. There are still many challenges, including the following:

- Determining accurate location, height (floor), and orientation of the phone is one of the biggest technical issues to solve. There are several commercial solutions combining different sensors in the phones to get an estimate, but the results need to be tested. Another potential solution is to ask users to input their location and floor number after the earthquake into a questionnaire.
- This test placed the phones in an ideal location, i.e. on the top floor at the corner of the building with the phones lying on the floor. In reality, only a few phones may satisfy these requirements. Testing of phones on different floors, in different locations, and on different surfaces (desk, couch etc.) is also necessary.
- Tests in different building types will also provide more insight into the types of information we can extract from the phone records for different buildings.
- Making use of ambient vibrations of the building and nearby small earthquakes to extract a building's characteristic parameters will expand the capability of SHM from the phones. More work is needed to determine if ambient noise recordings could be recovered from phones and the lower limits of earthquake shaking that can provide useful recordings.

These challenges illustrate that much research is needed before we can have a fully operational crowdsourced structural health monitoring system. However, the initial tests shown here illustrate the promise of using smartphones for structural health monitoring of buildings, and provide the basis for this future development of MyShake functionality.

5.8 Supplementary material

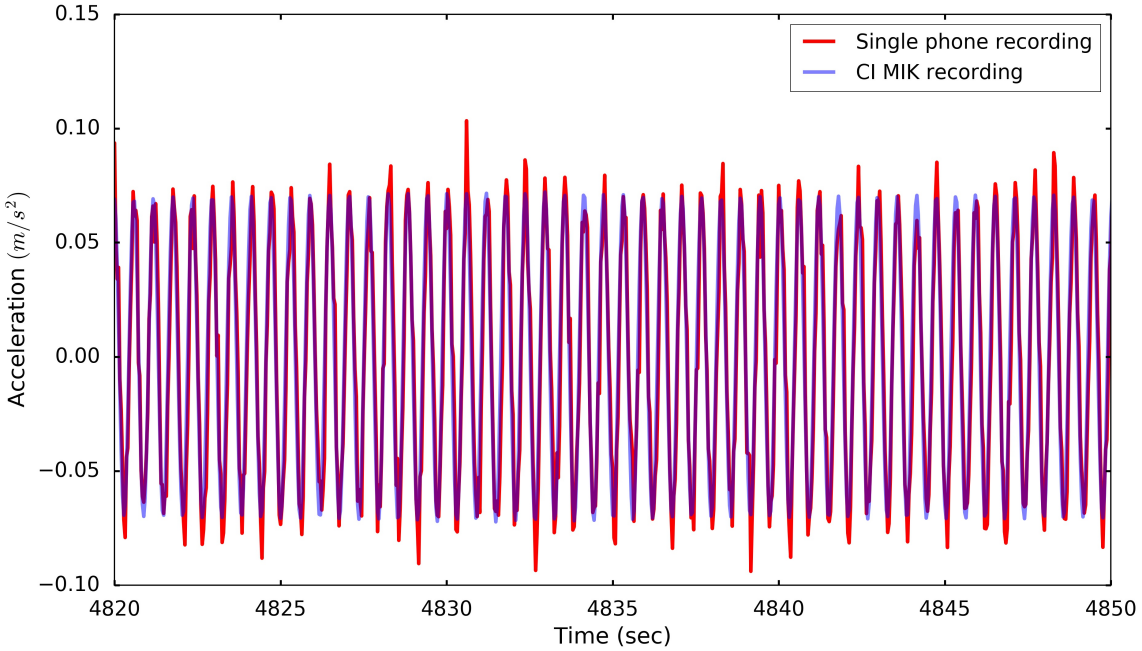


Figure 5.8: Acceleration record comparison for a single phone and the MIK station for time range 4820 s to 4850s (see **Figure 5.2**a for reference). The red recording is from a single phone (Samsung Galaxy S4) and the blue recording is from the MIK station. The phase alignment is excellent; the peak amplitudes show some deviation from the MIK recording. Also note that the phones and the MIK station are not in the same location; see the main text for their locations. The color version of this figure is available only in the electronic edition.

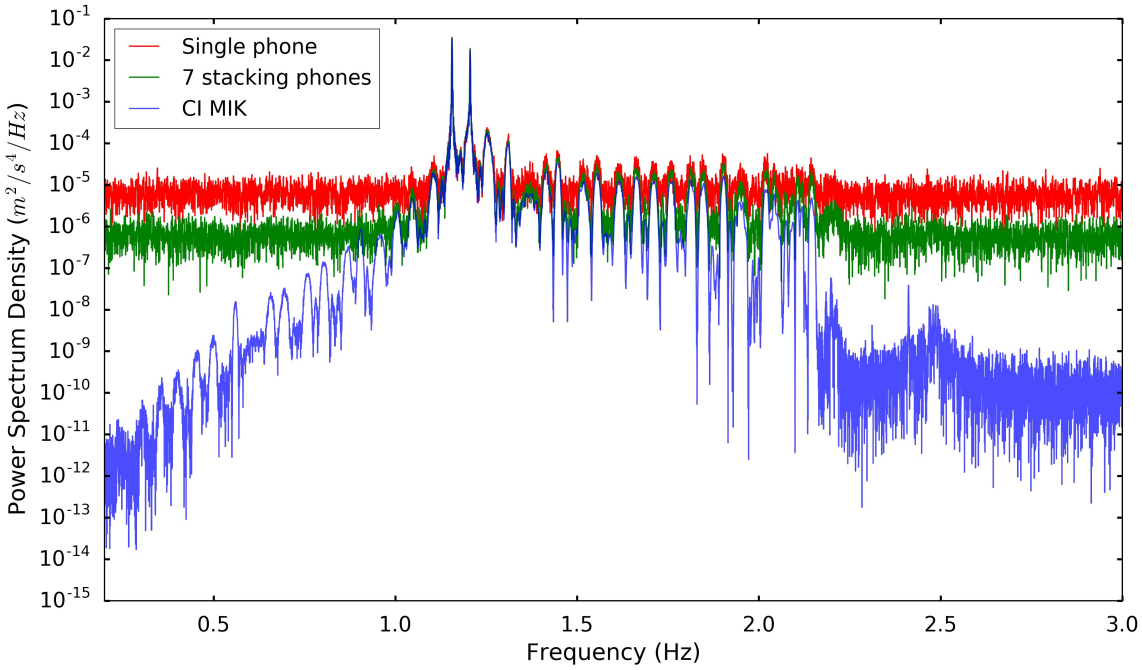


Figure 5.9: Spectrum comparisons for the E-W direction where the fundamental mode of the building is at 1.15 Hz and the fundamental torsion mode is at 2.35 Hz. Blue is spectrum for the Episensor (CI MIK), red is spectrum for a single phone (Samsung Galaxy S4) and the green spectrum is for the 7-phone stack. The color version of this figure is available only in the electronic edition.

Figure 5.10: A partial screenshot of the Millikan Library Cloudlet Display during the shaker tests. The display shows (from left to right, along the top): the instantaneous displacement of each sensor in the E-W direction (in 0.1 mm units); the displacements in the N-S direction; inter-story drift (a dimensionless measure of the difference in displacements between floors divided by floor height); a table of maximum displacements observed on each floor over the previous 10 seconds, 1 hour, and 1 day, for each measured axis. The lower left plot shows the displacement as a function of time for the sensor on the 9th floor. Note the sinusoidal oscillations in the N-S curve in the lower plot (red points) compared to the near-zero displacements in the E-W curve (blue points). The color version of this figure is available only in the electronic edition.

5.8.1 Structural health monitoring infrastructure

We have developed Cloud-based computational infrastructure that supports structural health monitoring by real-time analysis of acceleration data streaming from phone-based or more traditional sensors. The purpose of the infrastructure is to record and report detection signals and continuous observed measurements, during and immediately following shaking events such as earthquakes. Examples of this include rapid determination of peak ground motion measurements following earthquakes or explosions, maps and videos of ground and building shaking on small spatial scales, and monitoring the state-of-health of buildings for structural damage and global structural response.

A feature of the infrastructure is the provision of a local "Cloudlet," in practice a small, network-attached computer running specialized software, that is located in a building under study; it handles calculations of building properties appropriate for, and configured to, the building. The software running on the Cloudlet processor works with the pre-defined set of sensors which are located in the building. By scheduled and frequent contact with those sensors, the software obtains the latest sensor measurements and continuously re-computes the following quantities in real time: a) visual displays for on-site situational awareness parameters (websites, cell phone apps), b) location-specific frequency domain products, c) location-specific time domain products, and d) location-specific parameters. The Cloudlet processor also offers a set of HTTP endpoints, typically accessed over secure SSL connections. These endpoints deliver real-time Web browser displays that can be used by end-users such as building owners or managers to monitor the quantities listed above in pseudo or near real time.

We deployed a Cloudlet dedicated to the Millikan Library during the shaker tests described in this paper. In addition to the smartphone sensors and Episensors, ten MEMS accelerometers are permanently installed in the Millikan Library building, one per floor. Figure 5.10 shows part of a screen snapshot from the Cloudlet's browser display at a moment in time when the shaker was causing the building to be shaken along its N-S axis. It illustrates several calculated building products including current status of the building, and a table of the maximum displacements observed over the last 10 sec, 1 hour and 1 day. The Cloudlet hardware and software can be adopted by any building whose state of health is being monitoring by networked accelerometers such as cell phones.

5.9 Acknowledgements

This project is funded by Deutsche Telecom Silicon Valley Innovation Center, and by the Gordon and Betty Moore Foundation through Grant GBMF5230 to UC Berkeley. In the analysis of the data, Obspy [Beyreuther et al., 2010; Megies et al., 2011] and mtspec package were used (http://doi.org/10.5281/zenodo.321789); we thank the authors. We also thank Lucy Yin, Anthony Massari, Kenny Buyco and other students and staff for setting up the shaker tests.

Chapter 6 Conclusion

MyShake as a unique crowdsourcing project achieved much within its first 2 years. With the trial and errors, now we have much better understanding of the system, it can now handle the heavy load without any problems. But building the whole system to start to collect data is just the very beginning of the journey, exploring how we could utilize the data and making contributions to the earthquake community and society are our final goals. I hope the chapters in this thesis could provide you an overview of our approach to build this network - a combination of earth science with the data science under the umbrella of crowdsourcing. Even though this approach has its own limitations, it shows very promising results in helping us to understand earthquake physics as well as reduce the risk of the damages. The conclusion here will give you a summary of each chapter and the future work to make improvement of this system.

In Chapter 2 the procedures we took to prove the concept of the idea is presented. The noise floor tests help us understand the sensitivity of the sensors inside the phones – they could record earthquakes larger than M5 within 10 km, but with the newer sensors, we start to see the M3.5 shakings within certain frequency range. From these tests, it is clear that the sensors inside the phones could be used to record earthquakes even though the noise level is higher than the high-quality sensors. Besides, the shake table tests we did with the phones put freely on top it show the promising results that the phone could reproduce the shaking very well even they are not bolted on the table. In addition, it also shows the phones start to have clipped amplitude if they start to have relative motion to the shake table, but the frequency content can still preserved very well. After these tests, we started to build our prototype system and use it to collect the human activity data from the students from Berkeley and friends. 4 months data was collected to understand the human behaviors with the phones. We carefully investigated the difference between the shaking of human activities and that from the earthquakes (from shake table tests and the simulated data), and came up the Artificial Neural Network (ANN) approach to distinguish the earthquakes activities from the human activities. As we need to do this in real-time to detect the earthquakes, we only used 2-sec window to extract the features from both types of the shakings. We found out a combination of amplitude features with the frequency features could do a decent job by distinguishing the shakings. The trained ANN algorithm is implemented in the application that could be downloaded by the users. Once the ANN algorithm on the phone detected something like earthquake, it will send a short message data including time, location, and amplitude back to the server for further confirm for earthquakes. At the same time, it stores 5-min 3-component time series data and upload it to the server when it is connected to WIFI and power. A network detection algorithm running on the server checks if it is a real earthquake by aggregating the trigger data from multiple smartphones in a region. The whole system was implemented based on the design in this chapter and released to the public to start to collect data. There are many aspects could be improved on the design of the system, for example, currently, the ANN algorithm is designed using the same training data, and everyone download the application will use the same model. A better way to do this is to train the personalized model for each phone, so that the human specific behavior and the hardware on the phone will be all in the consideration of the better model. Two ways could be used to address this problem, first we collect all the human data on the server, and train a model for each phone

only use the human data from this phone and push the model back to the phone after training. But this centralized approach need to collect human data to the server and have the potential of privacy issues. An alternative is to use distributed training to train the model on the phone itself that eliminates the need to collect the data to the server. After training the model individually on the phone, we could collect the difference of the trained model to a reference model or some other summarized parameters to learn the specific of this particular phone. This better model trained for each phone will reduce the false detection right now that due to the underrepresented behavior of the students. On the server level, the current network detection algorithm is designed for small or medium network, and it will become really inefficient when we have millions of smartphones reporting the earthquake at the same time. There is a need to design new and efficient algorithms to cluster the triggers in time and space to address this problem.

After we released the application to the public, chapter 3 summaries the initial observations we obtained from this global smartphone seismic network. The trigger time of the phones during the earthquakes correlated with the P and S waves, the waveform comparisons of MyShake users and a nearby seismic station, the P-wave portion recorded on the waveforms, and so on all show the great promise of this global smartphone seismic network. With the excitement of collecting this great dataset, many potential fundamental earth science applications need to be tested and validated on the data. Some of the potentials of using the MyShake data are (1) detecting more earthquakes and make the event catalog more complete, especially at places where there are not so many seismic stations. The improve of the completeness of the catalog will enable us understand better of the tectonic environment, and the local fault systems; (2) this dense network data could potentially help us to study the earthquake source using some of the current existing methods like focal mechanisms, finite source modeling, back-projection and so on; (3) with a dense distributed seismic network, we are likely to collect more near-field data to help us understand and constrain the near-field deformation. It will also be really useful for the earthquake engineering community as well; (4) the high-resolution microzonation map is a possible product for our MyShake network, it will be extremely important to reduce the earthquake hazards with this finer-scale shaking map. There are many other applications with this network, but we should realize the difference of this smartphone seismic network with the traditional seismic network. First of all, this smartphone network is a dynamic network in space. Most people live in clusters in terms of space, therefore, our smartphone seismic network is high-biased towards these clusters. Besides, the configuration of the network is constantly moving and changing, due to the movements of the phone owners with time. This will be so different than the fixed seismic network we used decades in seismology. How could we take into account of this dynamically changing both in time and space into our detection of the earthquakes? To understand the nature of the network is extremely important for the further seismological applications we mentioned above. In addition, the recording instruments is not bolted on the ground but instead placed anywhere, desk, backpack, in the buildings, and so on, we need to calibrate the amplitudes recorded on the phones as well. All these challenges provide us an opportunity to combine data science with the earth science to address the problems.

Chapter 4 starts to introduce our exploration of using the data we collected to conduct seismological applications. Especially, the earthquakes characteristics, i.e. origin time, location and magnitude are very important in terms of evaluate the earthquakes. Therefore this chapter reports our ability to use the waveforms recorded on the smartphones to estimate these parameters. Using the grid-search approach, we could get a reasonable origin time and the location of the earthquakes. The magnitude of the earthquake could be estimated using the

standard local magnitude estimation used in California. But could we design better algorithms using machine learning to efficiently find out these parameters? Especially if we have a large and dense network, I can envision the machine learning based approach will give us faster solutions. Besides, the waveforms recorded by the smartphones might be clipped if the shaking is too large, could we recover the clipped signal by using the machine learning approaches? More interestingly, since the smartphones has the timing error associated with the data we recorded, but whether we could use the magic of large number of phones to cancel out the errors from multiple resources, and get a better estimation of the origin time and location. The same goes to the amplitudes of the waveforms we used to estimate the magnitude of the earthquake, could this large number of waveforms from a dense smartphones help reduce the magnitude estimation error? All these need to be tested when we have a large number of users and a dense network. Computationally, everything right now is processed on the cloud server using a centralized approach. We could image of the traffic from all the phones try to communicate with this central server when we have millions of smartphones in a region. With the emergent of the edge computing [*Shi and Dustdar*, 2016; *Satyanarayanan*, 2017], distributing the communication and computing to the edge nodes in our global smartphone seismic network could potentially address this problem.

Not only MyShake data will be useful to the seismological applications, but also beyond seismology as well. Chapter 5 discusses the potential of using MyShake data in structural health monitoring by doing a shaker test to shake a 9-story building. The data collected the smartphones shows that we could extract the fundamental frequencies of the buildings in different directions. This opens the door to use MyShake data to monitor the health state of the buildings before and after the earthquakes, and identify the potential damage of the buildings during the earthquakes. The advantage of using the smartphones to monitor the health state of the buildings is the ease of deploying these sensors. There is no need to deploy the sensors inside the buildings, since people are living in buildings. Therefore, we could easily set up a large-scale sensor network for the city level or beyond. But as discussed in this chapter, we still have many technical challenges to solve before we can enable the monitoring. Such as finding the accurate location for the phones in the buildings, especially which floor are important; the building types and heights will have a large effect on extracting the fundamental frequencies, we need to build a database for the observations; using ambient noise will enable more applications for the monitoring, but it also a much harder problem as the phones have much higher noise levels. Also, MyShake data could be used in other applications as well. For example the waveform recording showing in chapter 3 where we could see the human responses after the large earthquake shaking passed the location can be used to infer the post-earthquake response. The information will be important to understand how the users react to the earthquake and to design better practices to reduce injury. In addition, the location data of the users before and after the earthquakes could be used to show the population distribution and by overlaying the shaking intensity map, we could use the information to guide the rescue team after the earthquake. But all these different applications would benefit from a large number of users using MyShake.

MyShake is a project illustrates the power of combining data science with earth science, and there are still many outstanding questions in both fields that will interest data scientists as well as seismologists. We believe these chapters in this thesis just layout the beginning of this project, and in the next few years, we hope using the data driven approach from data science with the domain science knowledge, MyShake will continue to grow to provide important dataset and useful applications in various directions. One important part is missing in this thesis is the

earthquake early warning applications we talked briefly in chapter 2. Currently we are working on this functionality to issue the warnings to the public, and we hope this function will be soon available to the world.

Bibliography

Allen, R. (2011), Seconds before the big one., *Sci. Am.*, *304*(4), 74–79, doi:10.1038/scientificamerican0411-74.

Allen, R. M. (2007), Earthquake Hazard Mitigation: New Directions and Opportunities, in *Treatise on Geophysics*, vol. 4, pp. 593–625, Elsevier.

Allen, R. M. (2012a), Geophysics: Transforming earthquake detection?, *Science (80-.).*, *335*(6066), 297–298, doi:10.1126/science.1214650.

Allen, R. M. (2012b), Transforming Earthquake Detection?, *Science (80-.).*, *335*(6066), 297–298, doi:10.1126/science.1214650.

Allen, R. M., and H. Kanamori (2003), *The potential for earthquake early warning in southern California.*

Allen, R. M., H. Brown, M. Hellweg, O. Khainovski, P. Lombard, and D. Neuhauser (2009a), Real-time earthquake detection and hazard assessment by ElarmS across California, *Geophys. Res. Lett.*, *36*(5), doi:10.1029/2008GL036766.

Allen, R. M., P. Gasparini, O. Kamigaichi, and M. Bose (2009b), The Status of Earthquake Early Warning around the World: An Introductory Overview, *Seismol. Res. Lett.*, *80*(5), 682–693, doi:10.1785/gssrl.80.5.682.

Allen, R. V. (1978), Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.*, *68*(5), 1521–1532.

Anon (2011), Rebuilding seismology, *Nature*, *473*(7346), 146–148, doi:10.1038/473146a.

Atkinson, G. M., and D. J. Wald (2007), "Did You Feel It?" Intensity Data: A Surprisingly Good Measure of Earthquake Ground Motion, *Seismol. Res. Lett.*, *78*(3), 362–368, doi:10.1785/gssrl.78.3.362.

Bakun, W. H., and W. B. Joyner (1984), The ML scale in central California, *Bull. Seismol. Soc. Am.*, *74*(5), 1827–1843.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010), ObsPy: A Python Toolbox for Seismology, *Seismol. Res. Lett.*, *81*(3), 530–533, doi:10.1785/gssrl.81.3.530.

Bossu, R., S. Gilles, G. Mazet-Roux, F. Roussel, L. Frobert, and L. Kamb (2011), Flash sourcing, or rapid detection and characterization of earthquake effects through website traffic analysis, *Ann. Geophys.*, *54*(6), 716–727, doi:10.4401/ag-5265.

Bossu, R., M. Laurin, G. Mazet-Roux, F. Roussel, and R. Steed (2015), The Importance of Smartphones as Public Earthquake-Information Tools and Tools for the Rapid Engagement with Eyewitnesses: A Case Study of the 2015 Nepal Earthquake Sequence, *Seismol. Res. Lett.*, *86*(6), 1587–1592, doi:10.1785/0220150147.

Bossu, R., R. Steed, G. Mazet-Roux, F. Roussel, C. Etivant, L. Frobert, and S. Godey (2016), The Key Role of Eyewitnesses in Rapid Impact Assessment of Global Earthquakes, in *Earthquakes and Their Impact on Society*, pp. 601–618, Springer International Publishing, Cham.

Bossu, R., M. Landès, F. Roussel, and R. Steed (2018), Felt Reports for Rapid Mapping of Global

Earthquake Damage: The Doughnut Effect?, *Seismol. Res. Lett.*, *89*(1), 138–144, doi:10.1785/0220170129.

Bradford, S. C., J. F. Clinton, J. Favela, and T. H. Heaton (2004), *Results of Millikan Library Forced Vibration Testing*, Earthquake Engineering Research Laboratory.

BRADFORD, S. C., J. F. CLINTON, J. FAVELA, and T. H. HEATON (2004), *Results of Millikan Library Forced Vibration Testing*.

Brocher, T. M. et al. (2015), The Mw 6.0 24 August 2014 South Napa Earthquake, *Seismol. Res. Lett.*, *86*(2A), 309–326, doi:10.1785/0220150004.

Bürgmann, R., P. A. Rosen, and E. J. Fielding (2000), Synthetic Aperture Radar Interferometry to Measure Earth's Surface Topography and Its Deformation, *Annu. Rev. Earth Planet. Sci.*, *28*(1), 169–209, doi:10.1146/annurev.earth.28.1.169.

Cheng, M. H., M. D. Kohler, and T. H. Heaton (2015), Prediction of wave propagation in buildings using data from a single seismometer, *Bull. Seismol. Soc. Am.*, *105*(1), 107–119, doi:10.1785/0120140037.

Chung, A. I., C. Neighbors, A. Belmonte, M. Miller, H. H. Sepulveda, C. Christensen, R. S. Jakka, E. S. Cochran, and J. F. Lawrence (2011), The Quake-Catcher Network Rapid Aftershock Mobilization Program Following the 2010 M 8.8 Maule, Chile Earthquake, *Seismol. Res. Lett.*, *82*(4), 526–532, doi:10.1785/gssrl.82.4.526.

Clayton, R. W. et al. (2011), Community seismic network, *Ann. Geophys.*, *54*(6), 738–747, doi:10.4401/ag-5269.

Clayton, R. W., T. Heaton, M. Kohler, M. Chandy, R. Guy, and J. Bunn (2015), Community Seismic Network : A Dense Array to Sense Earthquake Strong Motion, *Seismol. Res. Lett.*, *86*(5), 1–10, doi:10.1785/0220150094.

Clinton, J. F., and T. H. Heaton (2002), Potential Advantages of a Strong-motion Velocity Meter over a Strong-motion Accelerometer, *Seismol. Res. Lett.*, *73*(3), 332–342, doi:10.1785/gssrl.73.3.332.

Clinton, J. F., S. C. Bradford, T. H. Heaton, and J. Favela (2006), The observed wander of the natural frequencies in a structure, *Bull. Seismol. Soc. Am.*, *96*(1), 237–257, doi:10.1785/0120050052.

Cochran, E., J. Lawrence, C. Christensen, and A. Chung (2009a), A novel strong-motion seismic network for community participation in earthquake monitoring, *IEEE Instrum. Meas. Mag.*, *12*(6), 8–15, doi:10.1109/MIM.2009.5338255.

Cochran, E., J. Lawrence, C. Christensen, and RS (2009b), The Quake-Catcher Network: Citizen Science Expanding Seismic Horizons, *Seismol. Res. Lett.*, *80*(1), 26–30, doi:10.1785/gssrl.80.1.26.

Colombelli, S., A. Zollo, G. Festa, and H. Kanamori (2012), Early magnitude and potential damage zone estimates for the great Mw 9 Tohoku-Oki earthquake, *Geophys. Res. Lett.*, *39*(22), doi:10.1029/2012GL053923.

Colombelli, S., R. M. Allen, and A. Zollo (2013), Application of real-time GPS to earthquake early warning in subduction and strike-slip environments, *J. Geophys. Res. Solid Earth*, *118*(7), 3448–3461, doi:10.1002/jgrb.50242.

D'Alessandro, A., and G. D'Anna (2013), Suitability of Low-Cost Three-Axis MEMS Accelerometers in Strong-Motion Seismology: Tests on the LIS331DLH (iPhone) Accelerometer, *Bull. Seismol. Soc. Am.*, *103*(5), 2906–2913, doi:10.1785/0120120287.

Dashti, S., J. Reilly, and J. Bray (2011), *iShake: Using Personal Devices to Deliver Rapid Semi-*

Qualitative Earthquake Shaking Information.

Dashti, S., J. D. Bray, J. Reilly, S. Glaser, and A. Bayen (2012), iShake : The Reliability of Phones as Seismic Sensors, *15th World Conf. Earthq. Eng.*

Dashti, S., J. D. Bray, J. Reilly, S. Glaser, A. Bayen, and E. Mari (2014), Evaluating the reliability of phones as seismic monitoring instruments, *Earthq. Spectra*, *30*(2), 721–742, doi:10.1193/091711EQS229M.

Donnellan, A., L. Grant Ludwig, J. W. Parker, J. B. Rundle, J. Wang, M. Pierce, G. Blewitt, and S. Hensley (2015), Potential for a large earthquake near Los Angeles inferred from the 2014 La Habra earthquake, *Earth Sp. Sci.*, *2*(9), 378–385, doi:10.1002/2015EA000113.

Dou, S. et al. (2017), Distributed Acoustic Sensing for Seismic Monitoring of The Near Surface : A Traffic-Noise Interferometry Example, *Sci. Rep.*, (April), 1–12, doi:10.1038/s41598-017-11986-4.

Duennebier, F. K., P. N. Anderson, and G. J. Fryer (1987), Azimuth determination of and from horizontal ocean bottom seismic sensors, *J. Geophys. Res.*, *92*(B5), 3567, doi:10.1029/JB092iB05p03567.

Earle, P. (2010), Earthquake Twitter, *Nat. Geosci.*, *3*(4), 221–222, doi:10.1038/ngeo832.

Earle, P., M. Guy, R. Buckmaster, C. Ostrum, S. Horvath, and A. Vaughan (2010), OMG Earthquake! Can Twitter Improve Earthquake Response?, *Seismol. Res. Lett.*, *81*(2), 246–251, doi:10.1785/gssrl.81.2.246.

Earle, P. S., D. C. Bowden, and M. Guy (2011), Twitter earthquake detection: Earthquake monitoring in a social world, *Ann. Geophys.*, *54*(6), 708–715, doi:10.4401/ag-5364.

Ekstrom, G., and R. W. Busby (2008), Measurements of Seismometer Orientation at USArray Transportable Array and Backbone Stations, *Seismol. Res. Lett.*, *79*(4), 554–561, doi:10.1785/gssrl.79.4.554.

Evans, J. R., R. M. Allen, A. I. Chung, E. S. Cochran, R. Guy, M. Hellweg, and J. F. Lawrence (2014), Performance of Several Low-Cost Accelerometers, *Seismol. Res. Lett.*, *85*(1), 147–158, doi:10.1785/0220130091.

Farrar, C. R., and K. Worden (2007), An introduction to structural health monitoring, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, *365*(1851), 303–15, doi:10.1098/rsta.2006.1928.

Faulkner, M., M. Olson, R. Chandy, J. Krause, K. M. Chandy, and A. Krause (2011), The next big one: Detecting earthquakes and other rare events from community-based sensors, *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sens. Networks*, 13–24.

Faulkner, M. et al. (2014), Community sense and response systems: Your Phone as Quake Detector, *Commun. ACM*, *57*(7), 66–75, doi:10.1145/2622633.

Feng, G., Z. Li, B. Xu, X. Shan, L. Zhang, and J. Zhu (2016), "Coseismic Deformation of the 2015 Mw?6.4 Pishan, China, Earthquake Estimated from Sentinel-1A and ALOS2 Data," *Seismol. Res. Lett.*, *87*(4), doi:10.1785/0220150264.

Finazzi, F. (2016), The Earthquake Network Project: Toward a Crowdsourced Smartphone-Based Earthquake Early Warning System, *Bull. Seismol. Soc. Am.*, *106*(3), doi:10.1785/0120150354.

Fleming, K., M. Picozzi, C. Milkereit, F. Kuhnlenz, B. Lichtblau, J. Fischer, C. Zulfikar, and O. Ozel (2009), The Self-organizing Seismic Early Warning Information Network (SOSEWIN), *Seismol. Res. Lett.*, *80*(5), 755–771, doi:10.1785/gssrl.80.5.755.

Grapenthin, R., I. A. Johanson, and R. M. Allen (2014), Operational real-time GPS-enhanced earthquake early warning, *J. Geophys. Res. Solid Earth*, *119*(10), 7944–7965, doi:10.1002/2014JB011400.

Havskov, J., and G. Alguacil (2015), *Instrumentation in earthquake seismology*.

HEATON, T. H. (1985), A Model for a Seismic Computerized Alert Network, *Science (80-.).*, *228*(4702), 987–990, doi:10.1126/science.228.4702.987.

Holzer, T. L., and J. C. Savage (2013), Global earthquake fatalities and population, *Earthq. Spectra*, *29*(1), 155–175, doi:10.1193/1.4000106.

Homeijer, B., D. Lazaroff, D. Milligan, R. Alley, J. Wu, M. Szepesi, B. Bicknell, Z. Zhang, R. G. Walmsley, and P. G. Hartwell (2011), Hewlett packard's seismic grade MEMS accelerometer, in *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 585–588.

Hsieh, C.-Y., Y. Wu, T.-L. Chin, K.-H. Kuo, D.-Y. Chen, K.-S. Wang, Y.-T. Chan, W.-Y. Chang, W.-S. Li, and S.-H. Ker (2014), Low Cost Seismic Network Practical Applications for Producing Quick Shaking Maps in Taiwan, *Terr. Atmos. Ocean. Sci.*, *25*(5), 617, doi:10.3319/TAO.2014.03.27.01(T).

Kanamori, H., P. Maechling, and E. Hauksson (1999), Continuous Monitoring of Ground-Motion Parameters, *Bull. Seismol. Soc. Am.*, *89*(1), 311–316, doi:10.1103/PhysRevE.82.066103.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, *122*(1), 108–124, doi:10.1111/j.1365-246X.1995.tb03540.x.

Kim, S., S. Pakzad, D. E. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon (2007), Health monitoring of civil infrastructures using wireless sensor networks, *Proc. 6th Int. Conf. Inf. Process. Sens. Networks*, 254–263, doi:10.1145/1236360.1236395.

Kohler, M. D., T. H. Heaton, and M.-H. Cheng (2013), The community seismic network and quake-catcher network: enabling structural health monitoring through instrumentation by community participants, *Proc. SPIE - Int. Soc. Opt. Eng.*, *8692*, 86923X, doi:10.1117/12.2010306.

Kohler, M. D., A. Massari, T. H. Heaton, H. Kanamori, E. Hauksson, R. Guy, R. W. Clayton, J. Bunn, and K. M. Chandy (2016), Downtown Los Angeles 52-Story High-Rise and Free-Field Response to an Oil Refinery Explosion, *Earthq. Spectra*, *32*(3), 1793–1820, doi:10.1193/062315EQS101M.

Kong, Q., and M. Zhao (2012), Evaluation of earthquake signal characteristics for early warning, *Earthq. Eng. Eng. Vib.*, *11*(3), 435–443, doi:10.1007/s11803-012-0133-1.

Kong, Q., Y.-W. Kwony, L. Schreierz, S. Allen, R. Allen, and J. Strauss (2015), Smartphone-based networks for earthquake detection, in *2015 15th International Conference on Innovations for Community Services (I4CS)*, pp. 1–8, IEEE.

Kong, Q., R. M. Allen, L. Schreier, and Y.-W. Kwon (2016a), MyShake: A smartphone seismic network for earthquake early warning and beyond, *Sci. Adv.*, *2*(2), e1501055–e1501055, doi:10.1126/sciadv.1501055.

Kong, Q., R. M. Allen, and L. Schreier (2016b), MyShake: Initial observations from a global smartphone seismic network, *Geophys. Res. Lett.*, *43*(18), 9588–9594, doi:10.1002/2016GL070955.

Kong, Q., R. M. Allen, M. D. Kohler, T. H. Heaton, and J. Bunn (2018), Structural Health Monitoring of Buildings Using Smartphone Sensors, *Seismol. Res. Lett.*, doi:10.1785/0220170111.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015), ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.*, *8*(1),

14003, doi:10.1088/1749-4699/8/1/014003.

Kuhn, M., and K. Johnson (2013), *Applied predictive modeling*.

Kuyuk, H. S., R. M. Allen, H. Brown, M. Hellweg, I. Henson, and D. Neuhauser (2014), Designing a network-based earthquake early warning algorithm for California: ElarmS-2, *Bull. Seismol. Soc. Am.*, *104*(1), 162–173, doi:10.1785/0120130146.

Langbein, J. et al. (2005), Preliminary Report on the 28 September 2004, M 6.0 Parkfield, California Earthquake, *Seismol. Res. Lett.*, *76*(1), 10–26, doi:10.1785/gssrl.76.1.10.

Larson, K. M. (2009), GPS seismology, *J. Geod.*, *83*(3–4), 227–233, doi:10.1007/s00190-008-0233-x.

Lawrence, J. F. et al. (2014), Rapid earthquake characterization using MEMS accelerometers and volunteer hosts following the M 7.2 Darfield, New Zealand, earthquake, *Bull. Seismol. Soc. Am.*, *104*(1), 184–192, doi:10.1785/0120120196.

Lee, W. H. K., and International Association of Seismology and Physics of the Earth's Interior. Committee on Education. (2003), *International handbook of earthquake and engineering seismology. Part B : project of the Committee on Education, International Association of Seismology and Physics of the Earth's Interior*, Academic Press.

Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, and J. B. Ajo-Franklin (2017), Fiber-Optic Network Observations of Earthquake Wavefields, *Geophys. Res. Lett.*, *44*(23), 11,792-11,799, doi:10.1002/2017GL075722.

Luetgert, J. H., J. R. Evans, J. Hamilton, C. R. Hutt, E. G. Jensen, and D. H. Oppenheimer (2009), NetQuakes - A new approach to urban strong-motion seismology, *Am. Geophys. Union, Fall Meet. 2009, Abstr. #S11B-1707*.

Luetgert, J. H., D. H. Oppenheimer, and J. Hamilton (2010), The NetQuakes Project - Research-quality Seismic Data Transmitted via the Internet from Citizen-hosted Instruments (Invited), *Am. Geophys. Union, Fall Meet. 2010, Abstr. #S51E-03*.

Lynch, J. P., and K. J. Loh (2006), A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring, *Shock Vib. Dig.*, *38*(2), 91–128, doi:10.1177/0583102406061499.

McNamara, D. E., and R. P. Buland (2004), Ambient Noise Levels in the Continental United States, *Bull. Seismol. Soc. Am.*, *94*(4), 1517–1527, doi:10.1785/012003001.

Megies, T., M. Beyreuther, R. Barsch, L. Krischer, and J. Wassermann (2011), ObsPy - what can it do for data centers and observatories?, *Ann. Geophys.*, *54*(1), 47–58, doi:10.4401/ag-4838.

Minson, S. E., B. a. Brooks, C. L. Glennie, J. R. Murray, J. O. Langbein, S. E. Owen, T. H. Heaton, R. a. Iannucci, and D. L. Hauser (2015), Crowdsourced earthquake early warning, *Sci. Adv.*, *1*(3), e1500036–e1500036, doi:10.1126/sciadv.1500036.

Moreno-Gomez, A., C. A. Perez-Ramirez, A. Dominguez-Gonzalez, M. Valtierra-Rodriguez, O. Chavez-Alegria, and J. P. Amezquita-Sanchez (2017), Sensors Used in Structural Health Monitoring, *Arch. Comput. Methods Eng.*, *0*(0), 1–18, doi:10.1007/s11831-017-9217-4.

Naito, S., H. Azuma, S. Senna, M. Yoshizaw, H. Nakamura, K. X. Hao, H. Fujiwara, Y. Hirayama, N. Yuki, and M. Yoshida (2013), Development and testing of a mobile application for recording and analyzing seismic data, *J. Disaster Res.*, *8*(5), 990–1000.

Olson, M., A. Liu, M. Faulkner, and K. Chandy (2011), Rapid Detection of Rare Geospatial Events: Earthquake Early Warning Applications, *Proc. 5th ACM Int. Conf. Distrib. event-based Syst.*, 89–

100, doi:10.1145/2002259.2002276.

Ozer, E., M. Q. Feng, and D. Feng (2015), Citizen sensors for SHM: Towards a crowdsourcing platform, *Sensors (Switzerland)*, *15*(6), 14591–14614, doi:10.3390/s150614591.

Paek, J., K. Chintalapudi, J. Caffrey, R. Govindan, and S. Masri (2005), A Wireless Sensor Network for Structural Health Monitoring: Performance and Experience, *Cent. Embed. Netw. Sens.*, 1–10, doi:10.1109/EMNETS.2005.1469093.

Prieto, G. A., R. L. Parker, and F. L. Vernon (2009), A Fortran 90 library for multitaper spectrum analysis, *Comput. Geosci.*, *35*(8), 1701–1710, doi:10.1016/j.cageo.2008.06.007.

Reilly, J., S. Dashti, M. Ervasti, J. D. Bray, S. D. Glaser, and A. M. Bayen (2013), Mobile phones as seismologic sensors: Automating data extraction for the ishake system, *IEEE Trans. Autom. Sci. Eng.*, *10*(2), 242–251, doi:10.1109/TASE.2013.2245121.

Sakaki, T., M. Okazaki, and Y. Matsuo (2010), Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, *Proc. 19th Int. Conf. World Wide Web*, 851–860, doi:10.1145/1772690.1772777.

Satyanarayanan, M. (2017), The Emergence of Edge Computing, *Computer (Long. Beach. Calif).*, *50*(1), 30–39, doi:10.1109/MC.2017.9.

Shi, W., and S. Dustdar (2016), The Promise of Edge Computing, *Computer (Long. Beach. Calif).*, *49*(5), 78–81, doi:10.1109/MC.2016.145.

Snieder, R., and S. Erdal (2006), Extracting the Building Response Using Seismic Interferometry : Theory and Application to the Millikan Library in Pasadena , California, , *96*(2), 586–598, doi:10.1785/0120050109.

Snoke, J. A. (2009), Traveltime Tables for iasp91 and ak135, *Seismol. Res. Lett.*, *80*(2), 260–262, doi:10.1785/gssrl.80.2.260.

Stachnik, J. C., A. F. Sheehan, D. W. Zietlow, Z. Yang, J. Collins, and A. Ferris (2012), Determination of New Zealand Ocean Bottom Seismometer Orientation via Rayleigh-Wave Polarization, *Seismol. Res. Lett.*, *83*(4), 704–713, doi:10.1785/0220110128.

Survey, U. S. G., and D. J. W. and J. W. Dewey (2005), Did You Feel It? Citizens Contribute to Earthquake Science, *Fact Sheet*, (March).

Wald, D. J., V. Quitoriano, B. Worden, M. Hopper, and J. W. Dewey (2011), USGS "Did You Feel It?" internet-based macroseismic intensity maps, *Ann. Geophys.*, *54*(6), 688–707, doi:10.4401/ag-5354.

Worden, C. B., M. C. Gerstenberger, D. A. Rhoades, and D. J. Wald (2012), Probabilistic relationships between ground-motion parameters and modified mercalli intensity in California, *Bull. Seismol. Soc. Am.*, *102*(1), 204–221, doi:10.1785/0120110156.

Wu, Y., W. Liang, H. Mittal, W. Chao, C. Lin, B. Huang, and C. Lin (2016), Performance of a Low-Cost Earthquake Early Warning System (P -Alert) during the 2016 M L 6.4 Meinong (Taiwan) Earthquake, *Seismol. Res. Lett.*, *87*(5), 1050–1059, doi:10.1785/0220160058.

Wu, Y.-M. (2015), Progress on Development of an Earthquake Early Warning System Using Low-Cost Sensors, *Pure Appl. Geophys.*, *172*(9), 2343–2351, doi:10.1007/s00024-014-0933-5.

Wu, Y.-M., and T.-L. Lin (2014), A Test of Earthquake Early Warning System Using Low Cost Accelerometer in Hualien, Taiwan, pp. 253–261, Springer Berlin Heidelberg.

Wu, Y.-M., D.-Y. Chen, T.-L. Lin, C.-Y. Hsieh, T.-L. Chin, W.-Y. Chang, W.-S. Li, and S.-H. Ker (2013), A High-Density Seismic Network for Earthquake Early Warning in Taiwan Based on Low Cost Sensors, *Seismol. Res. Lett.*, *84*(6), 1048–1054, doi:10.1785/0220130085.

Xu, N., A. Broad, and D. Estrin (2004), A Wireless Sensor Network For Structural Monitoring, *Proc. ACM Conf. Embed. Networked Sens. Syst.*

Yin, R.-C., Y.-M. Wu, and T.-Y. Hsu (2016), Application of the low-cost MEMS-type seismometer for structural health monitoring: A pre-study, in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1–5, IEEE.

Yu, Y., R. Han, X. Zhao, X. Mao, W. Hu, D. Jiao, M. Li, and J. Ou (2015), Initial Validation of Mobile-Structural Health Monitoring Method Using Smartphones, *Int. J. Distrib. Sens. Networks*, *11*(2), 274391, doi:10.1155/2015/274391.

Zhong, Z., P. Chen, and T. He (2011), On-demand time synchronization with predictable accuracy, in *Proceedings - IEEE INFOCOM*, pp. 2480–2488.