

Ultrasound Thermal Dose Detection with Active PZT element during Ablation Therapy

Yoonsu Kim

Johns Hopkins University
3400 N. Charles St.
Baltimore, MD 21218, USA
ykim99@jhu.edu

Jae Won Lee

Johns Hopkins University
3400 N. Charles St.
Baltimore, MD 21218, USA
jlee392@jhu.edu

Abstract

Thermal monitoring during ablation therapy is instrumental to preserve surrounding healthy tissues and determine the completeness of operation. We propose non-invasive thermal monitoring method using ultrasound. We introduces external Active Echo (PZT element) to attach to the ablation tool, and process the ultrasound signal in real-time to provide thermal information. Machine learning methods such as SVM, Adaboost and K-NN method enable binary decision for certain temperature. In our experiment, the classification accuracy was up to 99% using SVM and large set ultrasound features. This paper shows feasibility of thermal monitoring using ultrasound with machine learning methods.

1 Introduction

Ablation therapy is widely used for removing tumor cells, but it is also important to preserve surrounding healthy tissues and completely eliminate malignant cells. Surgeons must be able to judge the extent of ablation successfulness. Thus, we propose new monitoring method using conventional ultrasound system with active PZT element. Increasing temperature in tissue may cause changes in ultrasound modalities such as speed of sound, frequency shift, attenuation and bandwidth. Although some of papers have already proved that changes in temperature induces small change on speed of sound, many other aspects are not known yet. Mehdi et al. have classified prostate cancer cell with RF time series ultrasound signal using SVM classifier.[1] Kidney

and liner tissue are successfully classified with ultrasound signal in 94 percent accuracy with random forest classification method.[2] In this paper, we will discuss which factors are mainly affected by temperature change, applying various machine learning classification methods and evaluate our implementation of classifiers. We trained with various features which are extracted from ultrasound signal. This paper is organized as follows. Section II gives some background knowledge about ultrasound and introduces how we collect data and how we select features. Section III explains how we implement SVM, Adaboost and KNN as our classification method and how we use those methods. Last section of this paper includes analysis and comparison on our results.

2 Background

Ultrasound signals are sound waves with high frequencies which surpass the audible limit of human hearing. Piezoelectricity is abbreviated as *PZT*. Generally, it is said to be that *PZT* is the electric charge that accumulates in certain solid materials in response to applied mechanical stress. In other words, piezoelectricity is equal to electricity resulting from pressure. Recently, external piezoelectric ultrasound sources are used for tracking and high accuracy calibration.

2.1 Data Collection

In order to collect sample data for our project, we used a PZT element source, driving board, ultrasound DAQ system, thermometer and optical camera as our apparatus. PZT element source converts an electrical charge to ultrasound. Ultrasound

DAQ system allow us to gather raw ultrasound signal stream before forming B-mode image. This system synchronously records all data driven from our apparatus. The PZT element driving system transmits ultrasound when it receives trigger, and the DAQ system collects simultaneously. Chicken breast was continuously heated during data collection. First, we collect necessary data by ablating a chicken breast at fixed position. Secondly, once ablation starts, we used ultrasound DAQ system to collect all ultrasound channel data and recorded ablated area and temperature by using a camera. Figure 6 in Appendix section demonstrates our data collection. The ultrasound probe is connected to Ultra-Sonix DAQ system and it performs as an ultrasound receiver sensor. The ultrasound signal is transmitted from Active PZT element.

We collected 380 data and each of them was acquired in every 2 seconds during ablation. Range of temperature was from 23°C to 50°C and ablated radius for the last data was approximately 5mm. Since ultrasound RF channel data and optical images are synchronized to ultrasound DAQ system and sorted by time sequence, we can label them with binary number $\{0, 1\}$ based on threshold temperature at the point of 5mm apart from ablation center. Later we can modify threshold to examine if there is any difference with results. Figure 7 and 8 in Appendix section demonstrate the ultrasound RF signal received by the probe. Horizontal axis denotes the element number from 1 to 128 and vertical axis denotes number of samples after the DAQ system receives signal from the trigger. We sampled with frequency of 40Mhz. The ultrasound data is stored in IGTL format.

2.2 Feature Extraction

The speed of sound in tissue changes as temperature changes. Since time of flight is inversely proportional to speed of sound, we segment time of flight to each of element. The segmentation method is crucial, because the speed of sound only changes up to 2 percent in theory, and the ablation area is respectively very limited to ultrasound travel path. Moreover, signal includes lots of noises from various sources such as DAQ device and other external EMIs, reflected ultrasound and inhomogeneous of tissue, etc. After ablation, we have found that

the time of flight only changes 10 samples(250ns) at maximum for a certain element. Moreover, the segmented time of flight information which is shown in Figure 9 in Appendix section is highly affected by noises. As shown in Figure 8 in appendix section, the signal at the edge elements have low SNR because ultrasound probe have sharp drop in directional sensitivity. This small change in time of flight and noise interruptions often cause misclassification to our system. We used relative time of flight to initial states as a feature to minimize dependency on active PZT element location.

Other features are shifts in frequency, attenuations and tail size. Figure 10 and 11 in Appendix section are examples of extracted feature with linear regression fitting model. Such linear regression method is used for extracting more features. We used linear regression model on both of time domain and frequency domain. The reason that we used linear regression model parameters is that extracting bandwidth or tail length is difficult, it may require manual segmentation work.[1] Since the parameters such as coefficient covariance matrix, RMSE, loglikelihood, SST values are related to signal peak, width, and tail length changes, we used theses parameters as a part of feature vectors.

Additionally, we used an attenuation factor which reflects maximum intensity. Ablation causes change in ultrasound propagation impedance at tissue in small extent. Attenuation can also become an excellent feature for classification. Each feature includes 128 channel data, so the total size of feature is $128 \times \text{number of parameters of ultrasound we will use}$. In this paper, we used 2048 size feature vector.

3 Method

We use MATLAB to implement three different machine learning classification methods such as Support Vector Machine(SVM), k-Nearest Neighbor and AdaBoost to analyze our collected data. Since ultrasound thermal detection with machine learning is not studied thoroughly yet, main purpose of this study is to find the best classifier for our data. Although MATLAB provides libraries for SVM, k-Nearest Neighbor, we implemented our own 3 classification codes to adapt some options such as step sizes or update methods. Each method gives dif-

ferent results and will later discuss comparison between each methods. This section provides general idea of each classification method and how we apply each of them to our project.

3.1 Support Vector Machine

Support Vector Machines(SVM) are supervised learning that analyze data and recognize patterns, used for classification and regression analysis. SVM uses a flexible representation of the class boundaries and implements automatic complexity control to reduce overfitting. SVM is widely used for classification method, because it is easy to use, has good generalization performance and the same algorithm solves various problems with small modifications. Support Vector Machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. In short, SVM is to maximize minimum distance from data to separator. We implemented hinge loss function and three different gradient update methods, sgd, rmsprop, and adagrad. Hinge loss is loss function used for training classifiers. This is used for "maximum-margin" classification, most notably for support vector machines (SVMs). The hinge loss of the prediction y is defined as

$$l(y) = \max(0, 1 - f(x) \cdot y)$$

First, choose an initial vector form of parameters ω and learning rate α and η . Then for every iteration, the $\Delta Q_i(\omega)$ is updated from derivative of loss function, and we update weight vector multiplied by learning rate.

$$\omega := \omega - \alpha \eta \Delta Q_i(\omega)$$

α is constant, we set as 1 for our application, and η decreases with different rate. For the first update method, stochastic gradient descent(SGD) relaxes learning rate for every iteration by using number of iterations. The second update method is Adagrad, which uses gradient sum over iteration. The last update method is root mean square error (RMSProp). RMSProp method is similar to adagrad, but it introduces a decay rate to adjust gradient sum. Our algorithm generates random variables within dataset index to select a data sample.

3.2 k-Nearest Neighbor

Nearest neighbor utilizes previously generated or recorded examples to perform classification or regression on new data points. We have chosen to use euclidean and hamming distance to compute the similarity between an example and the point. This algorithm leads to sharp transitions between classifications where points that are very similar may be separated into different classes. The neighbors of the most similar point are often examined and incorporated into the calculation to provide smoother transitions. For MATLAB, *pdist2* library computes euclidean and hamming distance between two points. Euclidean distance is an ordinary distance between two points. This gives length of line segment connecting two points.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Although heat propagation generates different degree of wavefronts from 128 active PZT elements, data itself is still very sensitive so that it does not provide notable difference between each others. Therefore since hamming distance gives the number of positions or percentage at which the corresponding points are different, for some ultrasound data, hamming distance can be more appropriate way of determining difference from wavefronts. Rather than computing exact distance between each wavefront, hamming way only counts how many wavefronts are different in entire dataset.

3.3 AdaBoost

AdaBoost is an another machine learning classification method. This can be used in conjunction with many other types of learning algorithms to improve their performance such as misclassification error rate. AdaBoost algorithm works as follows. There is a learning algorithm called weak learner. Combining all weak learners into a weighted sum demonstrates the final result of this classifier. AdaBoost is short way of saying 'Adaptive Boosting'. Since results from previous classifiers are reflected to subsequent weak learners, we can name this as adaptive boosting. However, AdaBoost is also sensitive to noisy data and outliers. For our implementation, we select the highest accuracy among all features by

ignoring the same feature that has already been chosen. This option can be easily modified for other applications. Then we update α with error rate of weak classifier.

$$\alpha_k \leftarrow \frac{1}{2} \ln[(1 - E_k)/E_k]$$

If error rate increases, then α will decrease. Then we can compute weight vector with following equation.

$$W_{k+1}(i) \leftarrow \frac{W_k(i)}{Z_k} \times \begin{cases} e^{-\alpha_k} \\ e^{\alpha_k} \end{cases}$$

$e^{-\alpha_k}$ if correctly classified or e^{α_k} if incorrectly misclassified. Now in the end we can find weak learner with corresponding weight vector. This weak learner will be added to hypothesis class. Next weak learner which will be added to our hypothesis class will be calculated based on updated weight vector. With those all weak learners, we can finally predict with given equation below.

$$H(x) = \text{sign}([\sum_{k=1}^{kmax} \alpha_k h_k(x)])$$

We have tested with different size of hypothesis class and we can conclude that as hypothesis class increases, misclassification error rate decreases.

4 Result

We provide three different results based on three different classification algorithms that we implemented. Later in this section, we discuss about how change in threshold temperature will yield different results and different combinations of features will produce various outcomes.

Figure 1 shows the results from experiments with our implementation of AdaBoost classification method. We tested our AdaBoost with different number of Hypothesis. As number of hypothesis increases, the accuracy also increases. We used combined feature of first peak time of flight, FFT regression parameters, signal regression parameters, and max peak strength. We used train set of odd indexed samples, and tested with even indexed samples. This feature set consists of 2048 features. We set temperature threshold at 40 °C, the number of samples labeled either 1 or -1 are equal to 190. We have large

feature vectors, but have relatively small hypothesis classes for classification. Therefore we implemented our AdaBoost slightly different from classic Adaboost algorithm by the way of not selecting the same feature again. This classifier classified ablation threshold with accuracy of 93% with 5 hypothesis classes, and the accuracy increases along with number of hypotheses classes. We have observed selected hypothesis classes. We found that it mainly select attenuation factors and some elements of time of flight which are affected by ablation center. These selected hypothesis classes match with our expectation of element, index from 40 to 60 which would have changed more than others.

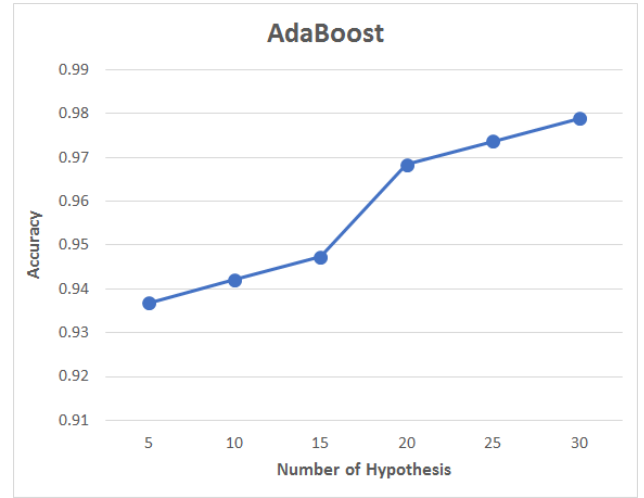


Figure 1: Result on AdaBoost algorithm

Figure 2 shows the result from our SVM classifier. We implemented three options of step size; Stochastic gradient descent, RMSProp and Adagrad method. Since SVM selects samples randomly, we performed 10 times and the graph shows the average result of accuracy. For our dataset, SGD shows the best result of more than 98% accuracy, and the other methods also performs above 90%. This might be caused by some outliers affected the accuracy of RMSProp and AdaGrad, because they use gradient sum information when they update. Therefore, the accuracy can be affected by the sample pick order. For example, outlier data added relatively large gradient to sum to minimize the learning rate for following samples.

We implemented k-NN with two different dis-

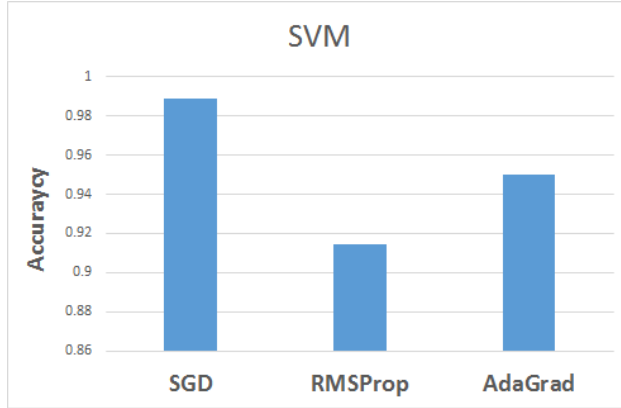


Figure 2: Result on SVM algorithm using different steps

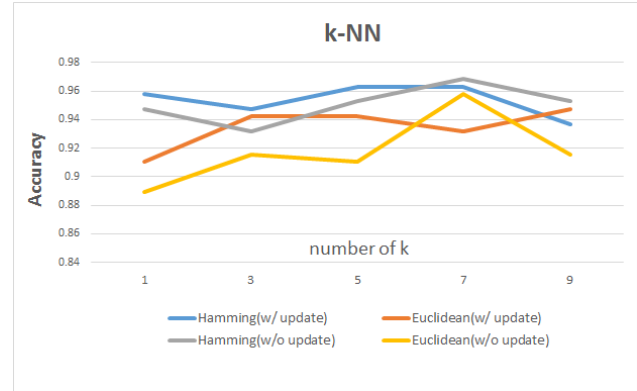


Figure 3: Result on KNN algorithm

tance calculation methods. We also implemented update option to add samples and labels to the classification set for later classification. We have 4 combinations with two distance measuring methods and updating options. The four line in Figure 3 shows the result with $k=(1,3,5,7,9)$. The k-NN with hamming distance classification produce worse result compared to that of Euclidean distance classification in major cases, but at some circumstances, hamming distance classification performs superior to Euclidean. In our dataset, the more important factor is change of some features over the numerical differences between samples. For example, after some heat propagations, change in number of ultrasound elements which have the time of flight and peak height increases. In some cases, the distance variation would be larger with hamming distance measurement. However, in our experiment, the Euclidean distance method shows better accuracy. A wrong prediction can induce inaccurate classification for following test with sample set. Thus, update option can decrease accuracy for some dataset. With larger number of neighbors, the classifier can increase the possibility of including differently labeled samples, so we can observe that the accuracy decreases when we pick 9 nearest neighbors. The optimal number of neighbors was 7 among single digit K set.

Our dataset includes ultrasound signal from 23 °C to 50 °C at the measurement point. Size of both train set and test set depends on threshold temperature. Therefore we performed cross validation by switching test set and training set. Since smaller size of

training set includes fewer outliers and larger training set fits better if it does not overfit the classification model, it shows better result for the case of 30 and 45 °C threshold. For our data, SVM shows better result over K-NN and Adaboost, because SVM often classifies over more flexible weight vector.



Figure 4: Cross validation

The purpose of this project is to define dominant

feature which determines the extent of ablation or temperature status. Result of time of flight features shows the lowest classification accuracy among all others. As we showed in data section, the segmented ultrasound time of flight is not quite accurate. This uncertainty can induce low accuracy in classification. Ultrasound segmentation issue has been problematic in ultrasound field. On the other hand, Adaboost still shows the best result on time of flight, because it would pick nicely segmented elements which received high peak signals. As we ablate tissues, the physical wave impedance of tissue changes. Therefore we can determine that attenuation factor can be the most dominant feature when monitoring temperature on tissues. Linear regression feature that combines FFT and signal model is also a promising candidate for thermal monitoring. The tail length of ultrasound can be longer and shorter according to the ultrasound impedance property changes in tissue.

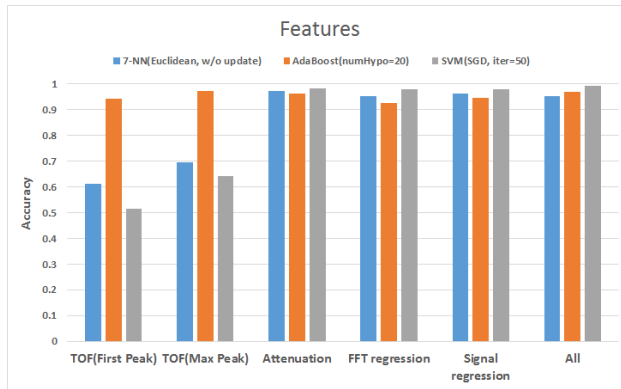


Figure 5: Select different features

We showed the feasibility of monitoring thermal effect on tissue during ablation using machine learning tool. We implemented three machine learning classification methods and evaluated them. The result shows that we can make binary decision in more than 90 percent accuracy by training with processed ultrasound features. In this experiment, the training set and test set is fixed at one position of PZT element. For better validation, it requires more dataset measured in various relative locations of PZT element relative to ultrasound probe in order to achieve more accurate outcomes. As a future work, we will collect ablation data with medical ablation tool with multiple PZT elements. By using improved ultra-

sound segmentation method, we expect that we can achieve more accuracy with smaller feature sets.

5 Comparison to proposal

- Our original objective is to implement SVM and compare with MATLAB SVM tools. However we have switched our gear towards implementing more classification algorithms such as k-Nearest Neighbor and AdaBoost, and SVM. We wrote every algorithms from scratch. We applied non-machine learning library such as distance measurement and histogram functions.
- At first, we expect that time of flight would be dominant feature for our data. However, at the end, final results demonstrate that attenuation factor and linear regression parameters was the most influential feature among all others.

Acknowledgments

The experiment data collection equipment was supported by MUSiiC laboratory of LCSR at Computer Science at Johns Hopkins University.

References

- [1] Mehdi Moradi et al. *Augmenting Detection of Prostate Cancer in Transrectal Ultrasound Images using SVM and RF Time Series*
- [2] Fereshteh Aalamifar, et al., *Classification of Kidney and Liver Tissue Using Ultrasound Backscatter Data*

6 Appendix

6.1 Code Structures

- MainFunction.m: Runs feature extraction, classifiers
- adaboost.m: AdaBoost Implementation
- KNN.m KNN Implementation
- SVM.m: SVM Implementation
- plotFigure.m: Provide results and Plot graphs
- AccCounting.m: Calculate accuracy rate
- ErroCounting.m: Calculate error rate
- LibFuncs/LoadData folder: Read IGTL data

6.2 Additonal Pictures

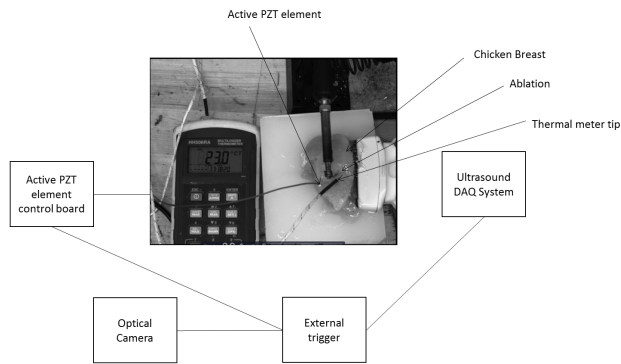


Figure 6: Setup for data collection

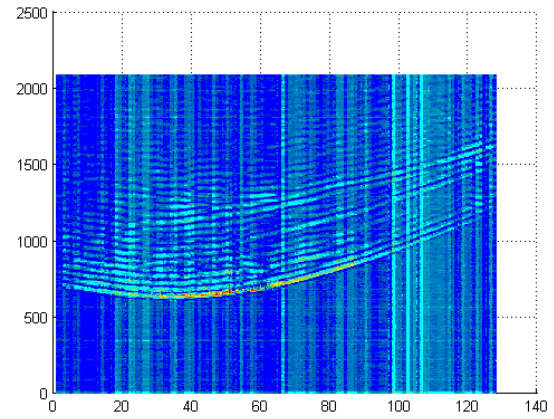


Figure 7: Ultrasound RF channel data of Active PZT element

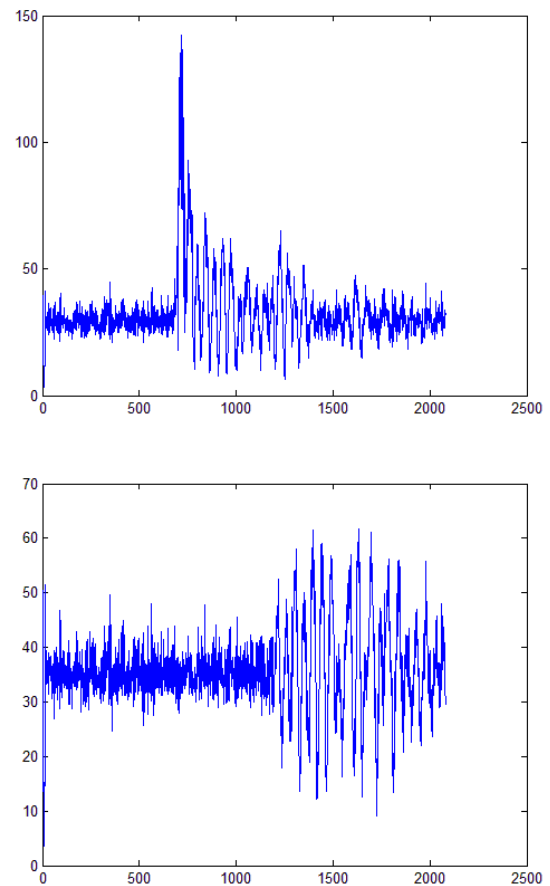


Figure 8: Ultrasound RF channel data collected from DAQ, top picture shows 64th element, bottom pictures shows 128th element signal

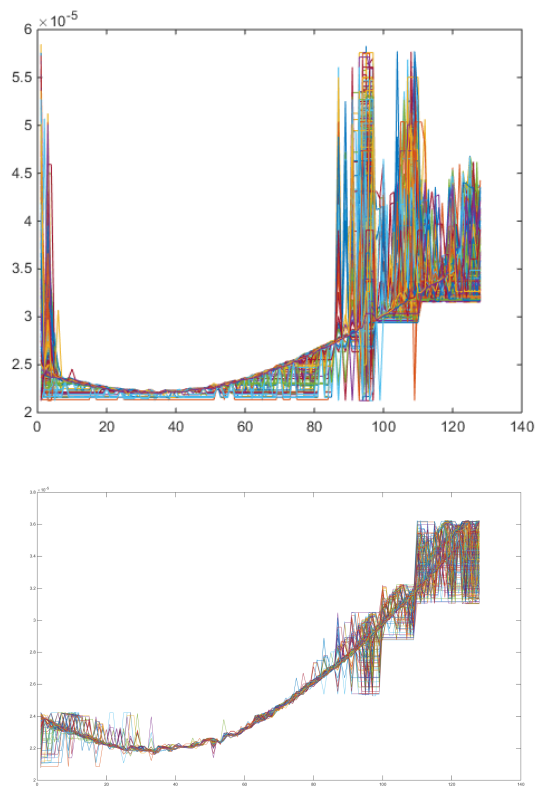


Figure 9: Time of Flight Segmentation, Top: First peak, Bottom: Maximum peak

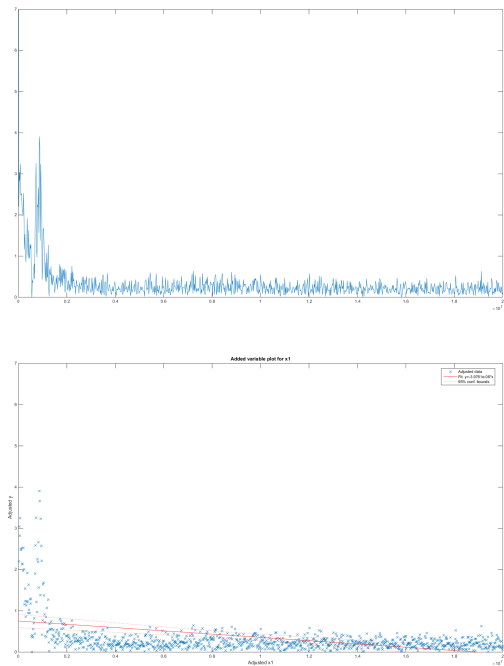


Figure 10: FFT of 45th element and its linear regression fitting

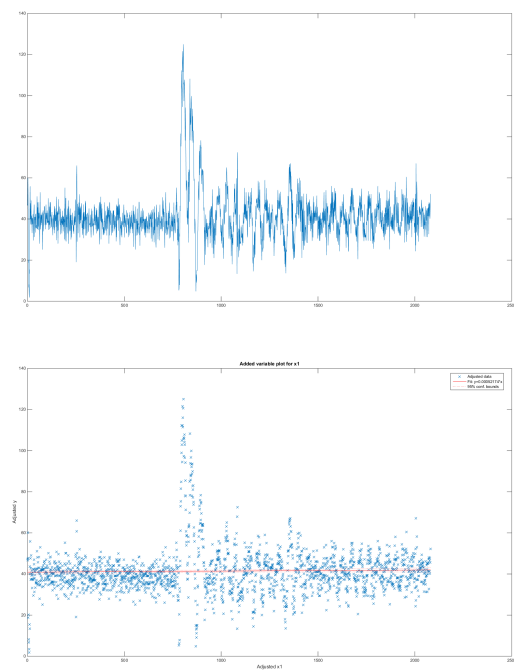


Figure 11: Signal of 45th element and its linear regression fitting