8th Quarterly Report – Public Page

Date of Report: *January 21, 2013*

Contract Number: *DTPH56-10-T-000021*

Prepared for: *Department of Transportation/Pipeline and Hazardous Materials Safety Administration*

Project Title: *Advanced Learning Algorithms for PIGPEN*

Prepared by: *Physical Sciences Inc.*

Contact Information: *Mickey Frish, frish@psicorp.com, (978) 738-8252*

For quarterly period ending: *December 31, 2012*

Public Page Section –

Work Completed During this Quarterly Period

Algorithm Development and Implementation

During this period, PSI prepared the self-learning algorithm for implementing real-time alarms. Previously, we reported analysis of data acquired during a parallel project at an Army test site and, using a trained classifier algorithm, post-processed data collected at a pipeline right-of-way test site. The algorithm has now trained itself based on data acquired after installation in a laboratory setting and subsequently identified abnormal activities that activate alarms.

Hardware Assembly and Field Test Preparations

As described in the previous quarters, cost-share partner Heath Consultants is building new sensor systems to be utilized in this project. The first sensor package was successfully installed and operated at the Heath Houston facility field test site. Delays in parts procurement delayed assembly of five additional sensor units, which were scheduled for assembly this quarter. Heath plans to test the six-sensor system at the Houston facility in February.

General Information required on all Public Quarterly Reports

Results and Conclusions:

<u>*Real-Time Algorithm Testing*</u>

Figure 1 illustrates the Self Learning Algorithm (SLA) flow chart for real-time operation; the SLA in a new environment undergoes a training phase before switching to the operational phase. In the training phase, the SLA passively acquires seismic data. Once all the training data has been acquired, the algorithm first extracts features from the signals. Then, it classifies (unsupervised) the signals using principal component analysis (for dimensionality reduction), k-means clustering (for clustering the reduced dimensions) and mahalanobis distance (to generate an activity index). This is followed by the creation of an identification (ID) database for signals. Each ID contains information about the population, activity index (fitted to a Gaussian) and date/time. As described in previous reports, signals are assigned IDs based on their dominant features. In this implementation, signals coming from different disturbances carry different IDs and can thus be discriminated.

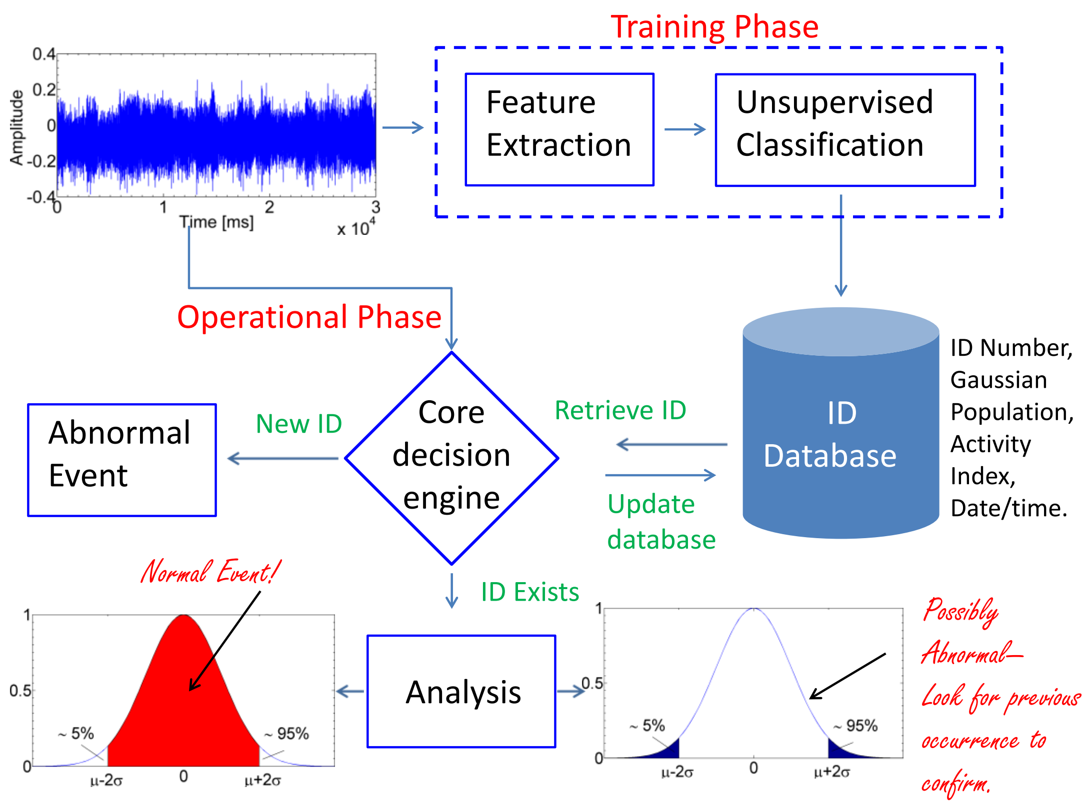

Figure 1. Flowchart of self-learning algorithm (SLA) operating in real-time.

In the operational phase, a core engine processes each incoming 30 second signal and decides whether or not the event is abnormal. Once a signal is received, the core engine determines its ID and looks in the ID database for this particular ID. If this is a new ID, then the event is flagged as abnormal. If the ID exists, the core engine looks at the Gaussian distribution of the population. If the activity index falls in the 5–95% confidence interval, the event is considered normal. If it falls outside this confidence interval, the event may be abnormal depending on whether previous occurrences exist. Also, the SLA determines whether the previous occurrences happened on the same day and around the same time.

For testing the algorithm in real time, a single seismic sensor was installed in an office at PSI's headquarters. The activity index plot for the background is shown in Figure 2. The term "background" refers to a so-called quiet environment where no major disturbances (such as humans walking, loud noises, etc…) are prominent. The two magenta lines represent the 5% and 95% confidence interval limits. Any event with activity index falling in between these lines (such as the red cross shown in Figure 2) is considered normal. Any events falling outside these lines are not in the confidence interval and are further processed by the core engine as briefly described above.

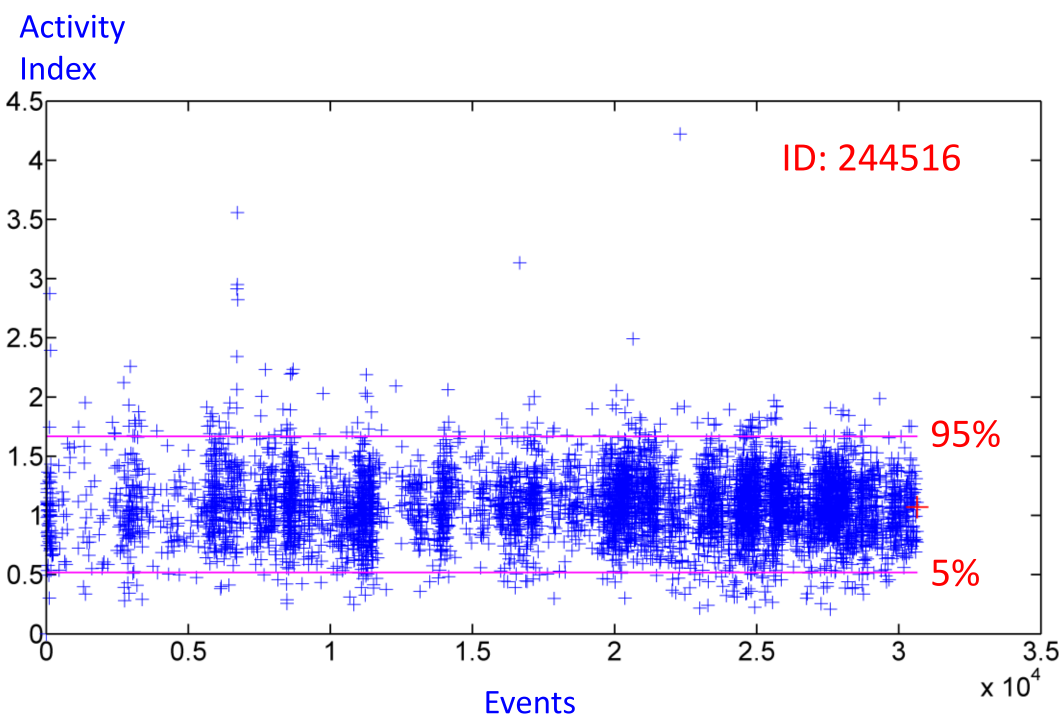

Figure 2. Background activity index plot.

The activity index plot of humans walking a few feet away from the sensor is shown in Figure 3. Verification that the ID assignment corresponds to walkers is done by visual observation. In Figure 3, we notice that the activity is higher during day time as compared to that of night time. i.e., activity index is greater than 1 (higher than the 95% confidence limit) during day time. This correlates with the fact that there is a lot more traffic during business hours. The red cross in the figure denotes a particular normal walking event.

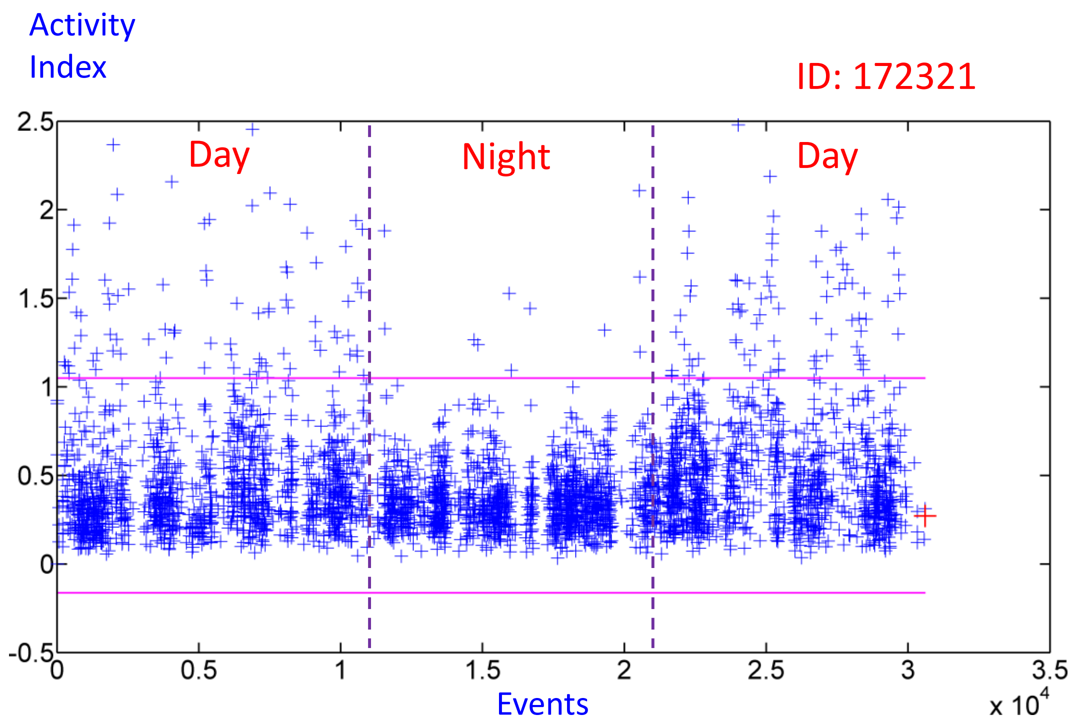

Figure 3. Activity index plot of humans walking a few feet away from the sensor.

To test the performance of the algorithm, it was challenged with normal and abnormal activities. For instance, heavy walking was purposely conducted several times. A representative example is shown in the activity index plot in Figure 4, where the red cross due to "heavy walking" appears above the 95% confidence interval. This triggers the SLA to search for events in the past having the "same" activity index (green dots in Figure 5), defined as a number plus or minus the standard deviation (std) divided by 10. In Figure 5, the blue lines delimit this activity index within ± std/10, and the cyan lines indicate the 5% and 95% confidence interval limits. The algorithm goes a step further and determines whether the past occurrences were around the same day and around the same time. In this case, the green dots were not circled and thus the event is flagged as abnormal.

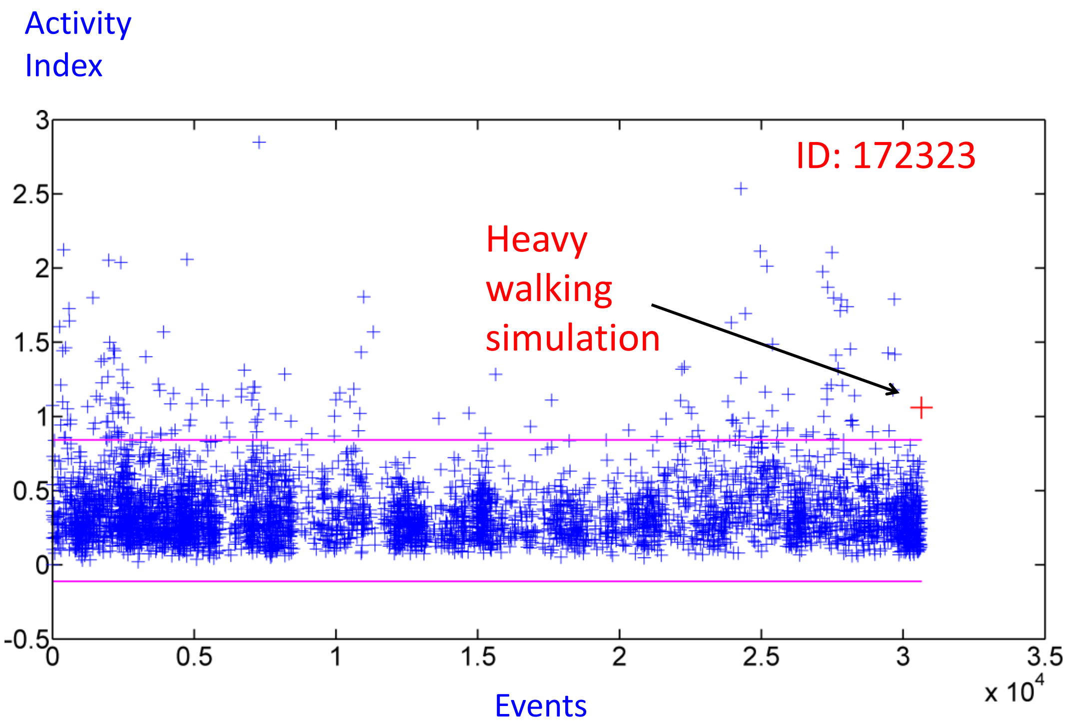

Figure 4. Heavy walking showing up on activity index plot for ID 172323.

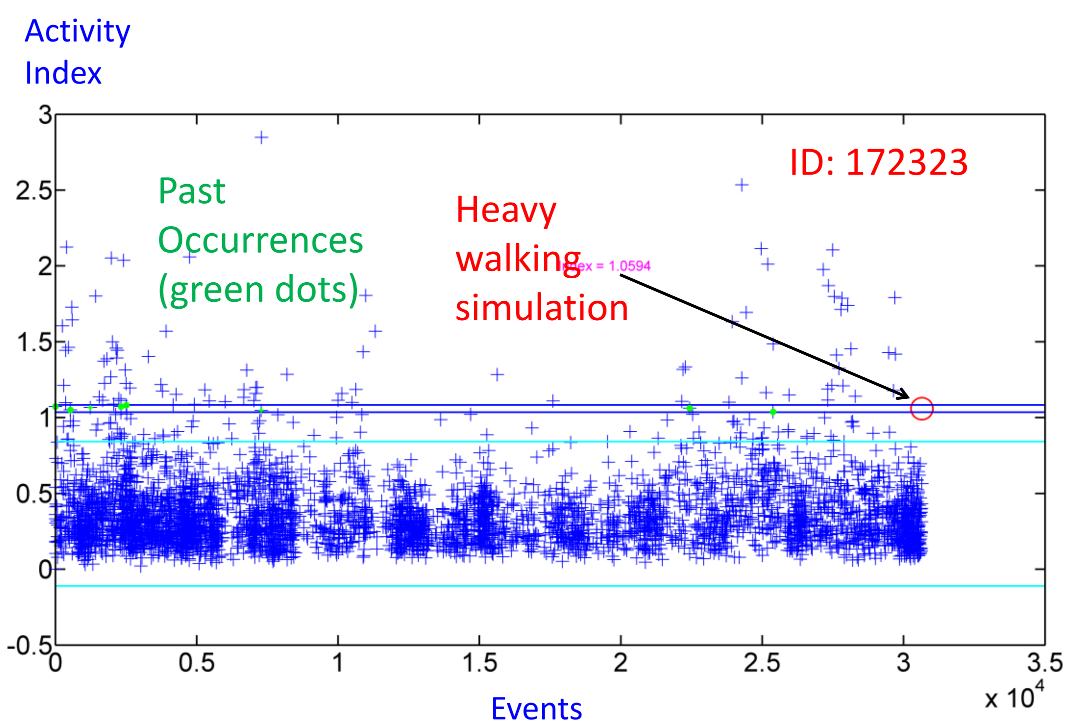

Figure 5. SLA searches for past occurrences of events with the same activity index.

Another purposeful challenge was a knock on the wall, creating a high activity index event with ID 244596, shown in Figure 6 (red cross). As seen from the activity index plot in Figure 6, the confidence interval region is not heavily populated as compared to the previous cases (human walking, background, etc…). The fact that the ID exists implies that wall disturbances (or related) do occur, but are rare compared to other above-mentioned disturbances. The activity index being above the upper confidence interval limit triggers the SLA to search for past occurrences. As shown in Figure 7, no past occurrences were found and the event was flagged as abnormal.

The SLA was also able to detect the sporadic "hissing" sound of the computer fan in the PC. This typically happens when the CPU is overloaded and the fan needs to work harder to cool down the CPU. As shown in Figure 8, the distribution mostly falls in the confidence interval, indicating that the loud noise has become routine. The not-so-populated distribution is also consistent with the sporadic nature of the event.

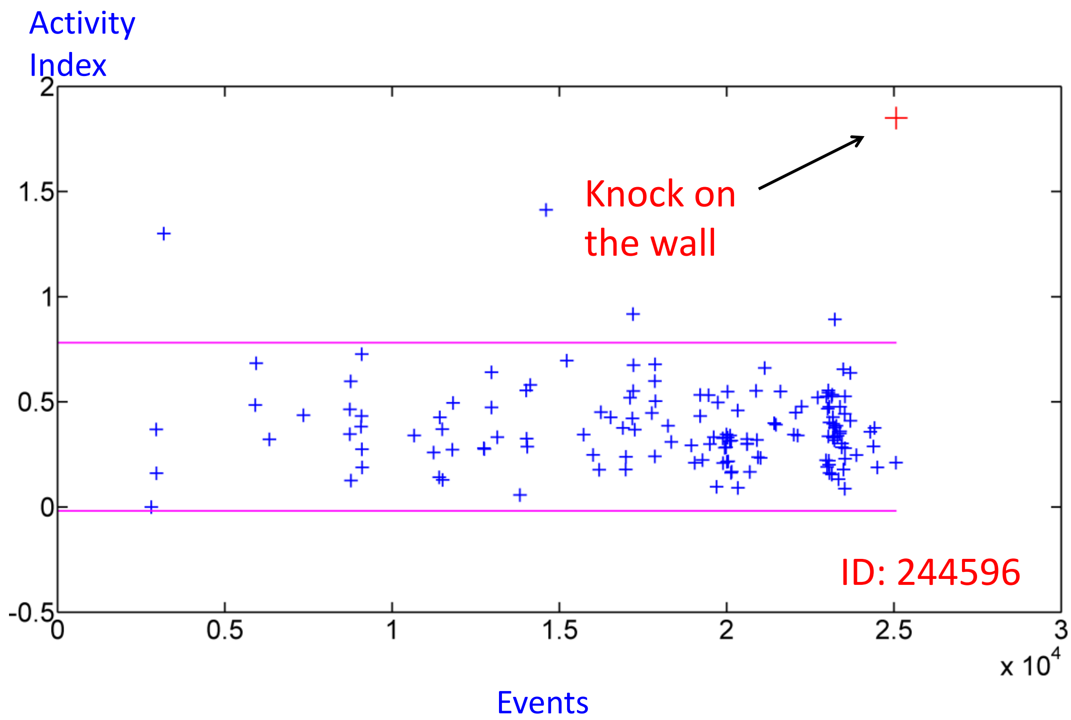

Figure 6. Real-life simulation of a knock on the wall.

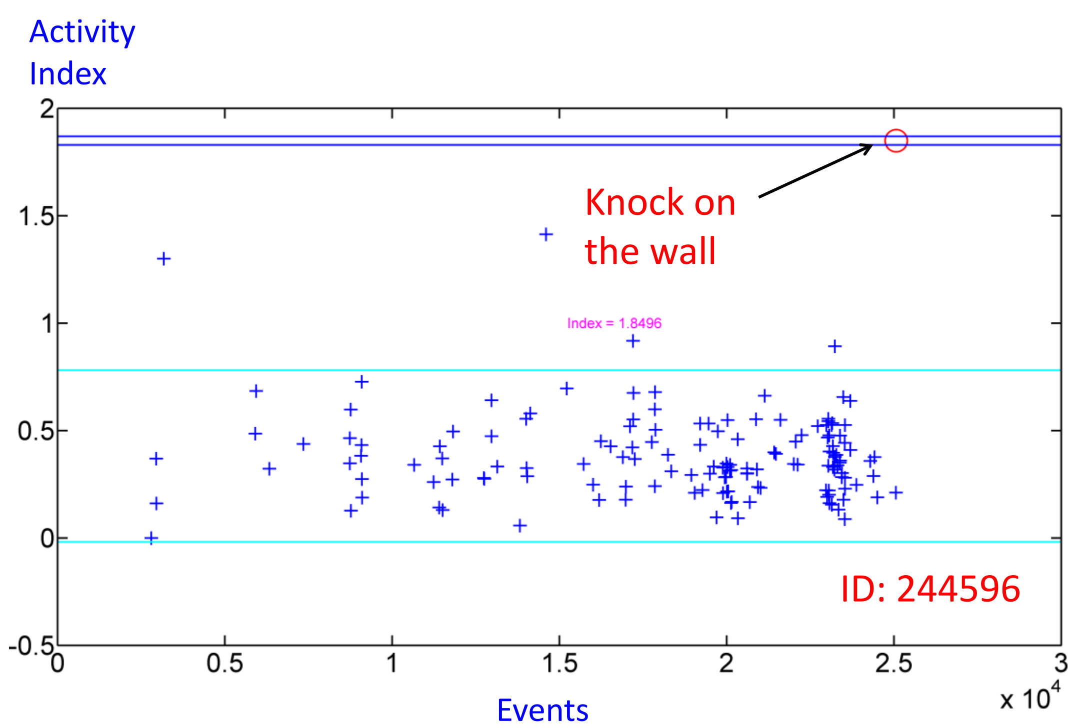

Figure 7. SLA does not see previous occurrences.

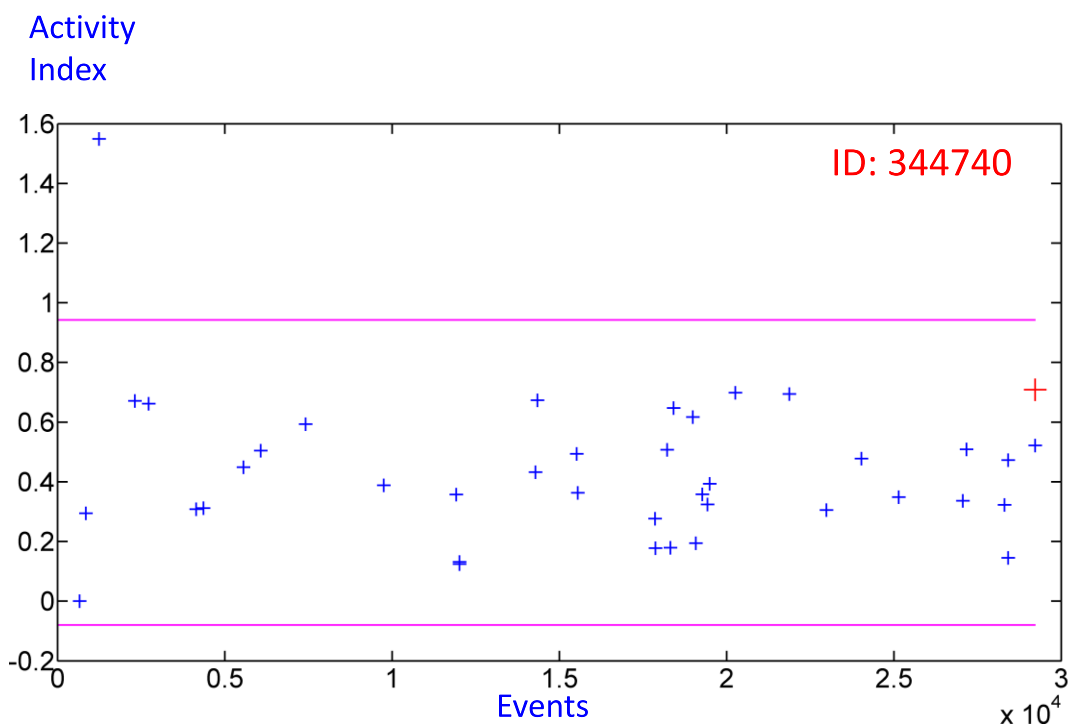

Figure 8. "Hissing" sound of fan when CPU is overloaded.

In summary, the SLA with its powerful ID assignment and robust core engine was able to discriminate between various seismic signals in real-time. The SLA was able to monitor specific disturbances, without supervised classification, over time and detecting changes in behavior (e.g. signal becoming louder or weaker). In addition, the core engine could locate previous occurrences and determine whether an event (outside the confidence interval) is repeating itself on the same day and time frame, or whether it's a new event.

Plans for Future Activity:

PSI, with cost share and product development partner Heath Consultants, are progressing with sensor assembly and field test planning. The initial field test using the new algorithms installed on new sensors is planned for February 2013.