A Signal Analysis Expert System for Signal Noise Reduction

Daniel T. Heinze

HRB Systems, Inc.
Science Park Road
State College, PA 16804

ABSTRACT
A Signal Analysis Expert System (SAES) has been developed to address the problem of noise editing pulse position modulation (PPM) signals in preparation for analysis. These signals historically required manual noise editing, or a combination of manual and semi-automatic editing techniques. Due to the nature of the signal, standard frequency domain noise reduction techniques are not applicable. The objectives for SAES were to provide a substantially improved tool for automatically noise editing the signal and to provide intelligent assistance to the signal analyst. To meet these goals, SAES employs a combination of statistical pattern recognition, image processing, and expert system technologies. The result is a tool which, for the signal of interest, reduces the noise editing workload on the analyst by approximately 80% over existing methods while retaining accuracy comparable to a human editor working manually.

INTRODUCTION
Signal analysis frequently involves a two-step process in which noise (and interfering signals) are first removed from the data, followed by an analysis process utilizing the edited data. The process of removing noise (i.e., noise-editing) can be time consuming and tedious, especially on those signals for which standard frequency domain filtering is inapplicable. Analysts trained to perform signal analysis typically regard the noise reduction step as a necessary evil which is a precursor to "getting on with the real job of signal analysis." Further, the noise reduction step has traditionally required a significant amount of analyst time and cannot be delegated to novice analysts because existing techniques of noise editing are either difficult to use, or do an incomplete job of editing. In response to these problems, we initiated an independent research and development project to develop a signal analysis expert system (SAES) aimed at semi-automatic signal noise reduction.[11,12] The objective of the SAES was to relieve highly trained analysts of the noise editing job.

The signal of interest for this project is pulse position modulated (PPM). The signal consists of a series of frames or time intervals with begin and end markers. During each frame a signal pulse arrives, and the position of the pulse within the frame determines the value for that particular sample. The major problem with this method of signal transmission is that interference causes not only a distortion of the signal pulses but also noise pulses which exceed the receiver threshold result in multiple false detections of the signal during each frame. Because each frame of the digitized signal has multiple values when noise is present, standard filtering processes are not useful in extracting the original signal. Figure 1 shows a simulation of a signal in a noisy environment. For signals with any significant level of noise, the typical noise editing method is to iteratively apply various editing methods and to then extract the signal from the remaining noise by manually tracing it across a display with a mouse or joystick. The entire process is laborious, time consuming, and the success of the available algorithms requires experience in selecting useful control parameters.

SAES OVERVIEW
SAES treats the signal of interest as a sparse binary image and extracts the original signal from the noise via a combined use of image processing, statistical pattern recognition, and expert system decision making. An analyst initiates the procedure by selecting a signal to edit. The signal is displayed in a window of the work station. The operator is queried via a pop-up menu to give a high level description of the frequency, amplitude variation, signal strength, and noise level characteristics of the signal. Once the signal features have been determined, they are input to an expert system which uses them to derive the correct parameters to use in the editing process. The editing process is performed by first convolving the image to determine which pulse in each frame is most likely the original signal. Figure 1 shows the result of one frame of the digitized signal with multiple values when noise is present. The analyst can then use the system to perform a minimum spanning tree on the remaining pulses, and a decision is made.

Figure 1: Even in low noise, a signal begins to lose definition.
for each node as to whether it possesses the characteristics of signal or noise. Nodes judged to be composed of noise are removed, and extended sections which appear to contain little or none of the original signal are marked for the analyst to review. Finally, data values are predicted for frames which have been left empty, and the original data are searched for values within a specified delta of the predicted value.

As an alternative to using an operator supplied signal description, SAES will operate in a fully automatic mode. In the automatic mode, SAES performs an initial edit using a generic set of parameters. The necessary signal features are extracted from the edited signal and adjustments to those feature descriptions are made based on a knowledge of the types of editing errors which would be expected given the type of features which have been identified. These features are passed to the expert system, and the noise editing process is applied to the original signal using the system derived signal description.

THE SAES CONVOLUTION FILTER

"Convolution is a classic image processing algorithm commonly used for spatial filtering and finding image features... The convolution operation replaces a pixel's value with the sum of that pixel's value and its neighbors, each weighted (multiplied) by a factor. The weighting factors are called the convolution kernel."[5] The size of the kernel and its weighting scheme depend on the objective of the filtering process. The selection of the correct kernel weighting scheme is essential to successfully convolving an image.[2,3,6,18]

The kernel weights used for SAES follow a normal (Gaussian) distribution with a standard deviation which varies according to the signal features, viz.

\[
\text{Weight} = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

where \( x \) is the mean distance between the point being convolved and the point being weighted and \( \sigma \) is the standard deviation of \( x \). In order to be distinguishable by human or machine, true data points must have a mean interpoint distance less than that between noise and data, or between noise and noise. Thus, over an extended range of frames, data points will be more closely aligned than noise points if a signal is discernible. For this reason, only the point in each frame which is closest to the point being convolved contributes to the weight of the point being convolved. This may be expressed as:

\[
w_{X,Y} = \sum_{j=1}^{c_x} \max(w_{X,j}, w_{Y,j}) + \sum_{j=1}^{c_x} \max(w_{X,j}, w_{Y,j})
\]

where:

- \( c_x \) is the convolved weight the point being convolved
- \( \sigma \) is the size-1 in \( x \) of the convolution kernel
- \( o \) is the assigned weight of the jth point inside the kernel in \( x \)

Where there is no data point in a column, or a noise point is chosen for some other reason, an error will result which will be corrected by a clustering algorithm. [8]

APPLICATION OF CLUSTERING TECHNIQUES

The second technique used in SAES involves a statistical classifier used to distinguish noise from data. An optimal classifier is possible if one knows the a priori probabilities and the class-conditional densities of the data.[7] This is almost never the case, hence many methods rely on generic design samples. For SAES, even design samples do not exist due to the irregular and varied nature of the input signal features. Algorithms which employ design samples are said to be supervised procedures. Those for which design samples are not used because they do not exist or because the patterns are too varied are unsupervised procedures.[4,14]

Standard statistical methods for unsupervised learning unfortunately produce inaccurate information when the data does not follow a normal distribution. However, "if the goal is to find subclasses, a more direct alternative is to use a clustering procedure. Roughly speaking, clustering procedures yield a data description in terms of clusters or groups of data points that possess strong internal similarities."[7]

The inability of clustering procedures to solve certain classes of problems and the connection between clustering and graph theory which seemed intuitively obvious led to Zahn's combination of the two.[13,14,20] Clustering algorithms form a group of operations which are classified as statistical area processes in which the image space is treated as a population and points are separated into clusters or "gestalts" on the basis of statistical measures on selected features. Zahn's work in this area was prompted by the fact that there were image groupings for the discernment of which "no algorithm... was known although [they] posed no difficulty for human perception." His algorithms stress the importance of proximity and similarity in perceptual groupings of points. The algorithms are implemented using minimum spanning trees (MST).

If the pixels in the image space are treated as nodes of a minimum spanning tree, clusters can be separated on the basis of edge weights. Two gestalt cluster theorems are of importance here:

If \( S \) denotes the nodes of \( G \) and \( C \) is a nonempty subset of \( S \) with the property that \( p(P,Q) < p(C,S-C) \) for all partitions \( (P,Q) \) of \( C \), then the restriction of any MST to the nodes of \( C \) forms a connected subtree of the MST.

If \( T \) is an MST for graph \( G \) and \( X, Y \) are two nodes of \( G \), then the unique path in \( T \) from \( X \) to \( Y \) is a minimax path from \( X \) to \( Y \).[20]

Put more simply (with some loss of completeness and accuracy), the first theorem says that the subtrees of the MST are clusters. The second says that if the MST is truly constructed, the nodes within a cluster will be connected to each other to form a subtree (i.e. they will not be connected to nodes outside the cluster other than being connected in such a way as to form a subtree of the MST).

Once the nodes are clustered, SAES classifies the clusters as data or noise on the basis of size, density, orientation, distance from other clusters, etc. and the noise clusters are pruned from the tree.
APPLICATION OF EXPERT SYSTEMS

The processes described here are deterministic. Given a raw signal and control parameters, the results are the same for every iteration of the noise editing algorithms. The problem in applying the algorithms comes in properly matching the control parameters to the raw signal. The algorithms could iterate through all the possible parameters, but the determination of which edited signal is optimal would be as or more complex a problem than producing the output. Also the time required for such an exercise is impractical. SAES addresses the problem with the use of an expert system to select the driving parameters.

The algorithms employed by SAES require eight driving parameters.\[8,11,12\] A human operator, capable of varying each parameter independently and without knowledge interrelated, and how the parameters relate to the characteristics of the unedited signal. To gain proficiency with the algorithms, the operator would need a detailed understanding of the algorithms, the algorithms in use, rules were developed for the selection of parameter settings and in what instances was there cross coupling between the operational descriptions and the system measurements in relation to parameter settings. Rules were then developed which produced parameter factors. Parameter factors are numerical values which when factored with the appropriate system measurements and a default parameter values produce the correct parameter settings for a representative set of signal conditions. Statistical routines were added to the system to maintain records of the usage and relative performances of the various editing sessions. Combining experience with an understanding of the algorithms in use, rules were developed for the selection of parameters under varied combinations of signal characteristics.

SAES uses two types of rules. The first determines the appropriate parameters to use given a set of signal features.\[15\] The second determines if the system is capable of correctly editing a signal with those features. If it is determined that there may be segments of the signal which cannot be correctly edited, the operator is warned and may invoke a manual editor to do some "quick" pre-editing of the signal to improve the signal to noise ratio in signal segments which could be troublesome.

The rules which determined if any pre-editing is necessary are straightforward. Certain combinations of signal characteristics are indicators that a pre-edit is needed. For example, a high noise level combined with a high frequency signal could could prove troublesome. Pre-editing is called for if certain severe combinations are present or if a sufficient number of less severe signal features are present.

The rules for determining the correct algorithm parameters are more involved and form a hybrid decision system with the noise editing algorithms themselves. The hybrid design of the system is required for two reasons. Firstly, all of the information necessary to make appropriate parameter determinations is not available to the system operator on the basis of visual observation of the unedited signal. Secondly, calls to a rule based expert system during the editing process would be too expensive in an already computationally burdened environment.

The parameters employed by the editing algorithms are all numerical, but the information provided by the operator is non-numerical and fuzzy. Aside from being fuzzy, the categories of information which the operator can provide do not directly correlate to any of the parameters used by the editing algorithms. The task of the expert system is to convert fuzzy, non-numerical information which does not match the algorithm parameter categories into exact numerical information which can be directly combined with the measurements collected during the editing process in order to produce correct parameter assignments.

Having exercised SAES through many signal examples each using various parameter settings, three pieces of information were available:

1. The operator supplied description of each signal
2. The measurements derived during the edit process
3. The success measure for each set of parameters

For the purposes of knowledge engineering, this information took the place of an expert. Because the technique was new, there were in fact no experts in its use and so this extensive operational information was vital. Using this information, it was determined which operational descriptions and system measurements had any relation to parameter settings and in what instances was there cross coupling between the operational descriptions and the system measurements in relation to parameter settings. Rules were then developed which produced parameter factors. Parameter factors are numerical values which when factored with the appropriate system measurements and a default parameter values produce the correct parameter settings for editing a particular signal segment. There are a few more than fifty rules for this task, but all of them have multiple conditions and results. The following example is a very simple illustration of this method:

\[\begin{align*}
&\text{o Set } \text{delta to } x \text{ (assign the default parameter)} \\
&\text{o If signal is high frequency then set the delta factor to } 1 \\
&\text{o Set } s \text{ to the measured standard deviation of MST arcs} \\
&\text{o Set prune algorithm } \text{delta parameter to: } \\
&\text{delta * delta factor * } s
\end{align*}\]

Using this method it is possible to combine high level operator observations and system measurements in a way which takes advantage of the expert systems' ability to produce exact numerical values based on composite, fuzzy descriptions and the editing algorithm's ability to combine this information with precise measurements without unnecessarily loading the processor. The load produced by calling the expert system during the edit process could be high due to the fact that several of the parameters are adjusted on a frame by frame basis during the edit.

The input to the expert system is made via a mouse sensitive menu (figure 2). The menu allows the operator to specify a range of values for each of five signal features - Frequency; Amplitude range; Signal strength; Range of discontinuities in the signal; Noise level. It is not necessary that the signal be uniform. The operator is allowed to specify more than one value for each of the signal features. For example, part of the signal may be high frequency and part of the signal may be low frequency; the noise level may range from high to low. All of the feature ranges are taken into account by the expert system, and parameters are optimized for the worst case scenario.

The advantages of this system are obvious. The operator need know nothing of the theory of the algorithms involved in the noise editing process. Tedium calculations and cross referencing are eliminated. Mistakes and retries
are reduced to a minimum. Because the signal characteristics which must be described are very high level and easily observed by even a novice, no special skills or expertise are required for doing noise editing on all but the most difficult signals. Even where some manual editing is required, it is not the tedious job of tracing of the signal with a cursor, but rather a simple matter of improving the signal to noise ratio by marking large areas of noise to be erased.

In the automatic feature extraction mode the human can be removed from the editing loop. Noise levels in the neighborhood of each frame are easily extracted. Extraction of features such as signal strength, frequency, amplitude levels, requires that the data be edited. This is accomplished by the use of a rough first edit followed by feature extraction. Features are extracted on a frame by frame basis and then passed through a smoothing algorithm to account for errors in the rough edit. These features are then input to the expert system and a final edit is performed.

During either manual or automatic editing, segments of the signal which SAES determines to have evidence of extensive signal loss or a strong interfering signal are marked and particularly called to the operator's attention. Once SAES has edited the data, the operator can review and approve it or make corrections as necessary.

**PERFORMANCE**

The SAES was tested and evaluated using both simulated and actual data. Because the signal and interference can both be accurately modeled, the use of simulations proved highly useful in the test and evaluation phase. Where cases of opportunity must be used for test and evaluation, it is likely that boundary conditions will not be adequately represented and the question of what represents truth cannot always be resolved. The methodology which we developed for test and evaluation uses simulated data and provides for thorough coverage of valid, boundary, and out-of-bounds cases. Additionally, an accurate representation of system performance as compared to humans can be derived.

Because SAES is a hybrid system which combines expert system technology with image processing and statistical pattern recognition, it was the operation of the entire system which was of concern, not just the operation of the expert system portion. Due to the manner in which the expert system was designed and based on our edits of a wide range of real and simulated signals, we knew that the expert system would produce good parameter factors if provided with the correct signal features. The primary question to be resolved by test and evaluation was: Could the operators quickly and accurately provide descriptions of the signal features which would enable the system to edit the signals with accuracy comparable to a human editor working manually? The operation of the system in the fully automatic mode was of concern as well.

SAES has been tested over a wide range of signal and noise characteristics. For illustrative purposes, four specific test cases on a particular signal are provided here.

1. 8 noise pulses per frame, no signal loss, no signal distortion.
2. 8 noise pulses per frame, 20% signal loss, signal distortion up to +/- 1% of bandwidth.
3. 20 noise pulses per frame, no signal loss, no signal distortion.
4. 20 noise pulses per frame, 20% signal loss, signal distortion up to +/- 1% of bandwidth, weak interference.
The performance goals established for each case were:

1. 100% accuracy after 1 attempt.
2. 100% accuracy after 1 attempt.
3. 90% accuracy after 2 attempts.
4. 90% accuracy after 2 attempts.

Testing was performed using simulated data for each of the four defined test categories. Using these data sets, tests were run to compare SAES to a human editor working manually. The results obtained by SAES and the human editor were compared against each other and against known truth. For actual data, there are no golden assets (known truth), and manual editing is considered the optimal method of noise removal. For this reason the test results represent the relative performance of SAES compared against manual editing with known truth serving as a calibration factor for the results. Because both SAES and the human editor were tested against known truth, it was possible for SAES to be measurably superior to the human editor. Such cases are evidenced in performance scores in excess of 100% for SAES.

The test results against the four test categories are summarized below:

<table>
<thead>
<tr>
<th>Performance</th>
<th>Passes Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>106%</td>
<td>1 (figure 3)</td>
</tr>
<tr>
<td>103%</td>
<td>1</td>
</tr>
<tr>
<td>100%</td>
<td>1</td>
</tr>
<tr>
<td>96%</td>
<td>1 (figure 4)</td>
</tr>
</tbody>
</table>

The test results are essentially the same regardless of whether SAES was run with operator input to the expert system or in the fully automatic mode. The length of time required for an analyst to noise edit a signal is reduced, on average, by 80% as compared to editing the same signal with the usual combination of manual and semi-automatic methods which the analysts were accustomed to using. The performance goals established for the project have been surpassed in terms of functionality, accuracy, and reduction of operator interaction with the system.

SAES was originally implemented in LISP on a Symbolics 3650. To improve performance, the core processing functions of SAES have been transferred to a forty processor INMOS Transputer which performs noise editing at 5X to 1X real time signal reception rates.[8,19]

This increase in speed has opened the way for further investigations and enhancements directed at retrieving signals even in the presence of severe interference and distortion.

**SUMMARY**

We have successfully implemented and tested an expert system to aid the signal analyst in the task of editing noisy PPM signals. The system utilizes a combination of image processing methods, statistical pattern recognition techniques, and expert system technology to perform either semi-automatic or automatic noise editing and interference reduction. Tests against both simulated and actual data have demonstrated that the system is able to automatically remove noise as well as experienced human signal analysts in most cases. Implementation on a parallel processing Transputer system provides real time processing rates.

Figure 3: A clean test type 1 signal is easily edited by SAES.
Figure 4: Even a badly corrupted test type 4 signal can be cleaned of false detections.

REFERENCES


