Theses - Daytona Beach   Dissertations and Theses

6-2006

Filtering of Acoustic Emission Data Through Principal Frequency Component Extraction

Justin O. Karl
Embry-Riddle Aeronautical University - Daytona Beach

Follow this and additional works at: https://commons.erau.edu/db-theses

Part of the Aerospace Engineering Commons

FILTERING OF ACOUSTIC EMISSION DATA THROUGH PRINCIPAL FREQUENCY COMPONENT EXTRACTION

By

Justin O. Karl

A Thesis submitted to the Graduate Studies Office

in partial fulfillment of the requirements for the degree

of Master of Science in Aerospace Engineering

Embry-Riddle Aeronautical University

Daytona Beach, Florida

June 2006

UMI Number: EP32100

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.   Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted.   Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
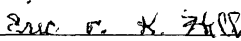
UMI®

FILTERING OF ACOUSTIC EMISSION DATA THROUGH PRINCIPAL FREQUENCY COMPONENT EXTRACTION

by

Justin O. Karl

This thesis was prepared under the direction of the candidate's thesis committee chairman, Dr. Eric v. K. Hill, Department of Aerospace Engineering, and has been approved by the members of his thesis committee. It was submitted to the Department of Aerospace Engineering and was accepted in partial fulfillment of the requirements for the degree of Master of Science in Aerospace Engineering.

THESIS COMMITTEE

Dr. Eric v. K. Hill
Chairman

Dr. Eric R. Perrell
Member

Dr. Seenithamby Sivasundaram
Member

Graduate Program Coordinator, MSAE

Department Chair, Aerospace Engineering

Date

i

ACKNOWLEDGEMENTS

A number of people have worked countless hours in support of me as a student to make this thesis possible Firstly, I would like to thank my parents, James and Pamela, as well as my brother Lake, for constant attention to my academic ventures and encouragement along the way I would like to thank my fiancée, Tiffany, for her commitments and support to me on both the emotional and intellectual levels

I would like to give thanks to Dr Eric Hill for providing me with the opportunity as well as the means and guidance necessary to have developed this far as an engineer and researcher To Dis Perrell and Sivasundaram I owe thanks for additional support and advice concerning the completion of this thesis

I would like to thank my fellow students Seth-Andrew Dion and Nicholas Spivey for their hours and hours of time spent working on analyzing nondestructive testing data on a number of projects, including this one Without their help, this thesis would never have developed to such an advanced stage

ABSTRACT

Author·   Justin O. Karl

Title·   Filtering of Acoustic Emission Data Through Principal Frequency
              Component Extraction

Institution   Embry-Riddle Aeronautical University, Daytona Beach, FL

Degree·   Master of Science in Aerospace Engineering

Year·   2006

Rapid editing of acoustic emission (AE) data is required in order to make real-time acoustic emission flaw growth systems a viable testing method for materials and setups that contain noisy signals. It was hypothesized that extracting major frequency components from the acoustic emission signal would therefore provide a representative acoustic signature of the major waveforms occurring due to defect growth This research has verified that the aforementioned filtering technique does, in fact, extract a representative signal from the composite and metal specimens utilized herein These findings were verified both through visual analysis of the data as well as the low error occurrence in backpropagation neural network predictions and good classification in self-organizing map type neural networks applied to the testing data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1.0 INTRODUCTION

1.1 Overview

The goal of this research was to verify major frequency extraction as a viable method for editing noisy acoustic emission data. Acoustic emission filtering of this type was given a number of requirements that were to be met in order to consider the filtering technique a success. Listed below are the requirements that were strictly adhered to during filter development.

- Filter must offer significant data reduction -- number of acoustic emission data points (hits) after filter is to be no more than 20% of the original number of hits.

- Filter must remove frequencies in the audible and near-audible ranges (< 50kHz) as well as excessively high frequencies in the megahertz range (> 2MHz).

- Filter must remove frequencies that commonly carry mechanical noise and multiple-hit data (low frequencies).

- Filter must identify frequencies that, when common AE plots are made, the plots closely resemble the ideal shapes and features shown in theoretically correct AE plots.

- Filter must identify frequencies that, when analyzed with a neural network or similar mathematical modeling system, produce low (<5%) error in prediction-type networks and clear patterns in classification-type networks.

The goals of the filter mentioned above were met after several iterations None of the filter iterations were far from meeting all of the requirements, and therefore it was never considered that the filtering mechanism might be logically or theoretically invalid in nature A complete overview of filter development will be covered in the section on filter theory

1.2 Specimen and Test Types

The data for this research initially came only from composite specimens, but were later expanded to include metallic specimens in order to test the overall validity of the filtering technique, since logically it should work for all types of material systems

The composite specimens used in this research were a set of 15-inch diameter graphite/epoxy over-wrapped pressure vessels manufactured in 1997 for use as test subjects at the NASA Marshall Space Flight Center (MSFC) These composite over-wrapped pressure vessels (COPVs) were comprised of an unknown graphite filament and an unknown epoxy matrix type, wrapped over a thin aluminum liner The COPVs were cured in different manners, including rotisserie and non-rotisserie cures Rotisserie cures are designed to ensure even matrix material distribution during cure The COPVs were artificially damaged in three different ways to simulate damage that might occur during storage, shipping, and installation into various spacecraft systems The damage included blunt and sharp impacts of varying energies, and lacerated hoop fibers Also, some of the COPVs were left undamaged

The data from the COPV specimens were taken during a number of proof tests of the pressure vessels Pressurization schemes and rates were varied in an effort to simulate usage in different types of systems In addition to varying pressurization schemes and rates, some of the COPVs were tested at ambient temperature while others were tested at cryogenic temperatures It should be noted that a large array of tests were being performed at once, thus, acquisition of acoustic emission data did not appear to be of primary concern Moreover, the acoustic emission system was not set up properly it therefore captured a large amount of low-level noise and multiple hit data This of course was another positive aspect for this filtering mechanism evaluation, because the noisy data presented a more difficult problem to solve

It would appear that the primary idea behind the testing at MSFC was to produce multivariate conditions that reflect realistic uses of the COPVs It was actually fortunate that this research was able to use this test data, because a success validates use of this type of filtering technique under widely varying conditions

The data from the metal specimens were somewhat less complicated The metallic specimen data came from a series of 7075-T6 aluminum tensile test specimens that were fatigued to failure in an MTS tensile test machine The specimens were tested and the data collected as part of a thesis project [Ibekwe, 2003] Subsequently, re-tests and re-filtering of the data were explored because the initial test results did not seem to correlate

with the available data It is possible that the data files were corrupted or partially lost

This mixture of known good data and partially lost or noisy data represented another

challenge for the filtering technique, which in this case was employed in a completely

different material system to verify its general applicability

1.3 Validation Methods

The validation of the filtering technique will be accomplished through the comparison of

the filtered and unfiltered data This comparison will be accomplished through well-

documented and historically utilized methods -- visual comparison of standard acoustic

emission plots and analysis of prediction-type as well as classification-type neural

network outputs

Visual comparison of acoustic emission plots first requires that the analyst understands

what standard acoustic emission plots look like The three standard plots that this study

used to examine the effectiveness of the filtering method were

- Duration vs counts,

- Amplitude vs duration, and

- Differential amplitude distribution

The first plot, duration vs counts, is a representation of the length of the acoustic

emission signal vs the number of times the waveform of the signal exceeds the threshold

value The plot ideally should be a scatter plot of data points that follow a highly linear

function This linear function represents the frequency of the dominant failure

mechanism in the material. Oftentimes the ideal case is not observed. Instead, the most ideal case one can hope for during analysis is a set of several frequencies indicating the primary frequencies of oscillation associated with the most dominant failure mechanisms in the material.
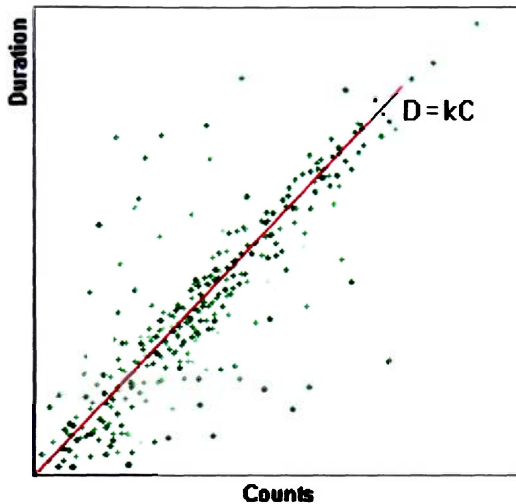
Figure 1: Idealized duration vs. counts plot

In the composite specimens, between one and five mechanisms are expected, while in the metal specimens one to three mechanisms are expected. In the case of this example plot, a linear trend line (which represents the mean frequency) is superimposed over the data. The constant, k, is used to describe the relationship between duration and counts, which again represents the mean or predominant frequency.

The second plot type, amplitude vs. duration, is another scatter plot that compares the maximum amplitude of an event with the length of the event. In this plot, it is common

for groupings of data points to occur that represent the different failure mechanisms in the material Citing the composite as an example, it is easy to understand that different mechanisms will have certain fixed relationships between amplitude and duration For instance, a fiber break is expected to have a high amplitude and a short duration, while a matrix crack would be expected to have a longer duration and small amplitude
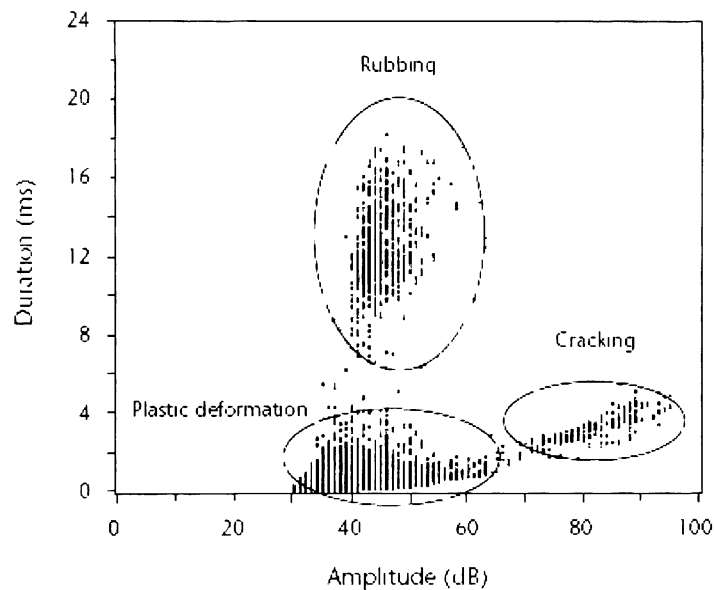
Figure 2 Sample duration vs amplitude plot [ASNT AE Handbook, 2005]

In Figure 2, three failure mechanisms are seen in the fatigue test of a bolted aluminum specimen with a notch in it to promote fatigue cracking The short duration/low-to-medium amplitude mechanism is plastic deformation ahead of the crack tip, and a longer duration/high amplitude mechanism represents the fatigue cracking itself The longest duration events in this plot are from the specimen rubbing within the bolted joints during cycling

The third plot, the differential amplitude distribution, often referred to as an "amplitude histogram," is a representation of how many acoustic emission hits occur at each amplitude between 0 and 100dB. This plot often displays a number of high points ("humps") and low points in the distribution, where these "humps" can be identified with certain failure mechanisms.

Figure 3: Idealized amplitude histogram

These "humps" in the histogram are well-defined in clean acoustic emission data but are generally blended together by noise in real world acoustic emission data sets. In 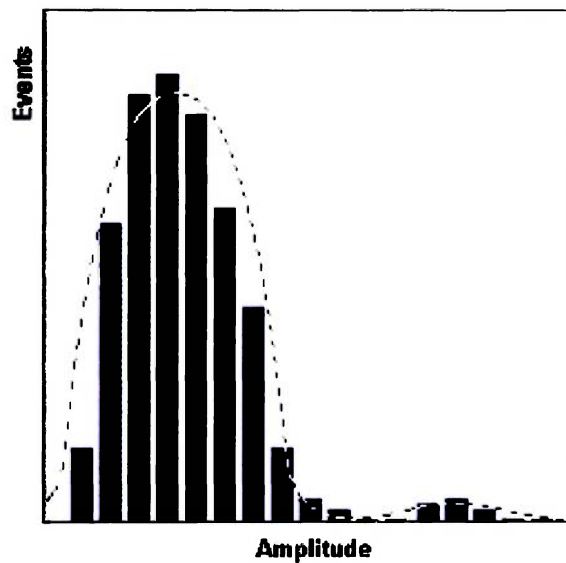this example it can be seen that there is a large failure mechanism at low amplitudes and a smaller failure mechanism at higher amplitudes.

Neural networks, a complicated set of iterative mathematical tools modeled after the mammalian brain, provide another way to validate acoustic emission data filters Self-organizing maps (SOMs) are a type of neural network that are used for classification The weight values of the network will cluster together after several iterations, leading to groupings that correspond to the failure mechanisms in the data SOMs commonly provide better classification than visual inspection of the duration vs amplitude plot and were used herein to show the improvement in the quality of both the composite and metallic specimen data after editing

Backpropagation, which are prediction-type neural networks, were also used to correlate COPV amplitude histograms with their corresponding results -- burst pressures After a number of amplitude histograms have been used for training, the backpropagation neural network can make predictions based on the historical data If the acoustic emission data are clean and well defined, the neural network can pick out patterns and correlate them to an output result within a small error margin Here it was desired to obtain a burst pressure to within 10% of what was observed during testing in order to consider that the neural network was making an accurate prediction, but manual filtering methods developed previously [Karl, Dion, Spivey, 2006] obtained results that were within 5% error Therefore, a 5% worst case error was set as the goal in this filtering exercise as well

2.0 THEORY

2.1 Acoustic Emission Overview

Acoustic emission, by definition, is the release of transient elastic waves propagating from flaw growth activity. This flaw growth activity has a number of sources on the microscopic as well as macroscopic levels. On the microscopic level, grain boundaries and microvoids can slip, coalesce, or slide against one another to release stress that is built up by deformation of the material. This reorganization of the atomic structure naturally brings the overall state of the system to a lower energy level, and the principle of conservation of energy dictates that energy must therefore be released. This release of energy occurs as a wave propagating in all directions from the source of the change in the material. This wave then undergoes reflections, mode conversions, and dispersion consistent with its path through the material -- continuing on until it is dampened out or transfers all or a part of its energy to another medium. In the case of AE instrumentation, this medium is the piezoelectric material within the transducer.

In metals, the flaw growth activity is primarily microscopic with the acoustic emission coming from sources like grain boundary slippage and rubbing, microvoids, microcracks, and particularly from plastic deformation associated with strain hardening of stress concentrated areas. In composite materials, the same types of microscopic emissions are detected, but they are often eclipsed by higher-energy activity that is primarily macroscopic in nature. The sources that will now be discussed assume a common

9

aerospace composite comprised of a high strength fiber embedded in a polymer matrix, arranged in layers of a laminate. The first source, matrix cracking of a macroscopic nature, occurs due to stress in excess of the ultimate strength of the matrix (which is typically much lower than the fiber material). Fiber breakage occurs at high stress when fibers begin to fail. Fiber pullouts occur when a broken fiber is pulled out of the matrix due to a lack of fiber/matrix adhesive strength. Delaminations occur in order to relieve shear stresses between layers of composite (lamina); and the final source, interlaminar rubbing, occurs when these separated lamina are brought back to their original positions after stress relief.

With a basic understanding of the possible sources, the next logical step is to gain an understanding of the operation of the acoustic emission data acquisition system. The propagation of the wave into the piezoelectric material in the AE transducer causes the piezoelectric material to vibrate, and the piezoelectric effect converts this vibration directly into a voltage. This fluctuating potential is picked up by a (positive) wire embedded in the piezoelectric material, with the reference (ground) being the metallic casing of the transducer itself, as seen in Figure 4.
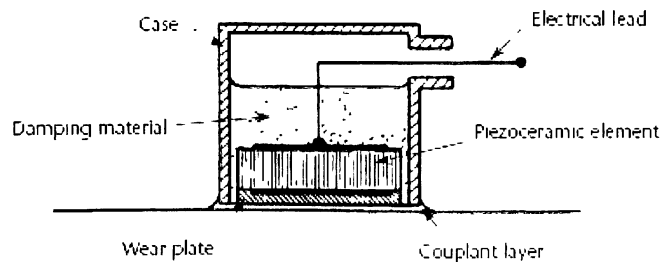

Figure 4: Wideband AE transducer (cross section) [ASNT AE Handbook, 2005]

Built into the casing of the AE transducer is a BNC connector, and a coaxial cable can be connected to the transducer to carry the signal from the piezoelectric material to the data acquisition system. Some transducers have a preamplifier built into the transducer casing while others do not. A standard acoustic emission acquisition system is shown in the following block diagram.
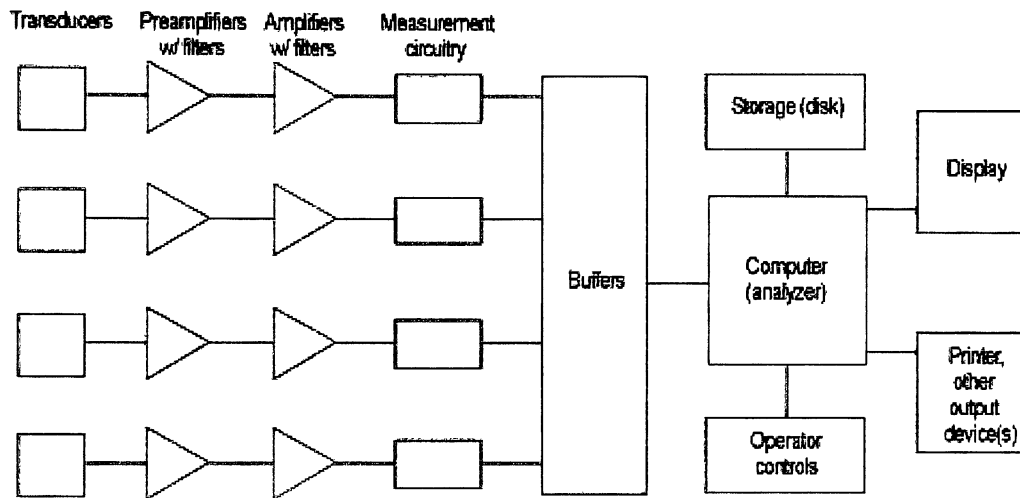

Figure 5: A standard AE system

After leaving the AE transducer, the signal first passes through a series of filters. Generally, these filters are band pass filters that are used to remove any electronic or mechanical noise in the signal. Electronic noise and mechanical noise are characterized by very high and very low frequencies, respectively. After this initial bandpass filtering, the signal, which even if pre-amplified is still very weak, passes through a variable amplifier. This amplifier has the effect of changing the signal from a maximum value in the range of microvolts up to the range of volts. Acoustic emission systems commonly

11

have an overall gain between 60 and 120dB from the original AE signal, depending upon if a preamplifier (commonly either 40dB or 60dB) is used.

After amplification, additional filtering may be used to further reduce the noise in the signal. This filtering is dependent upon the application and type of system being used. The portion of the signal remaining after amplification and filtering is passed on to a set of circuitry that calculates waveform parameters based on the signal. These waveform quantification parameters are the commonly referred to AE parameters that are used in this and many other research papers. Pictured in Figure 6 is an illustration of an arbitrary AE waveform and how the parameters are defined.
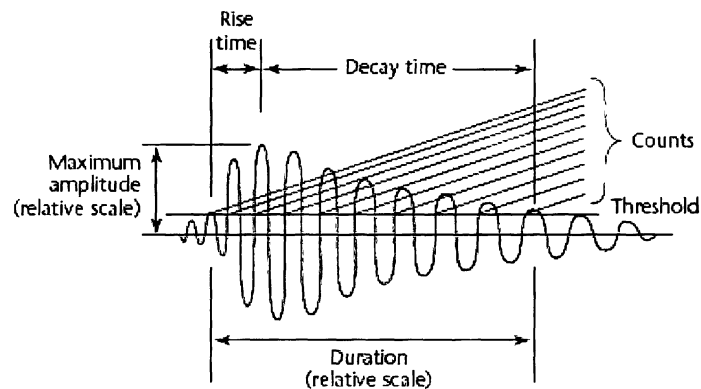

Figure 6: AE waveform parameters [ASNT AE Handbook, 2005]

A threshold is set by the user of the AE system to eliminate low-level noise. This simple technique instructs the system to ignore any signals that are not energetic enough to have an amplitude value above the threshold. A common threshold value for composite material testing is 60dB. Waveforms that do meet all of the filtering requirements and

are strong enough to exceed the threshold are characterized by the system into the AE parameters seen in Figure 6. Maximum amplitude (usually referred to as simply "amplitude") is a measure of the highest amplitude of the waveform. Rise time is the length of time (in microseconds) from the first crossing of the threshold to the maximum amplitude. Duration is the length of time from the first crossing of the threshold to the last crossing of the threshold (in microseconds), while counts is the number of times the waveform crosses the threshold. Energy counts (or commonly just "energy") is the computed value of the area under the entire rectified waveform that exceeds the threshold. This area is referred to as the MARSE (mean area under the rectified signal envelope) and is pictured in Figure 7.
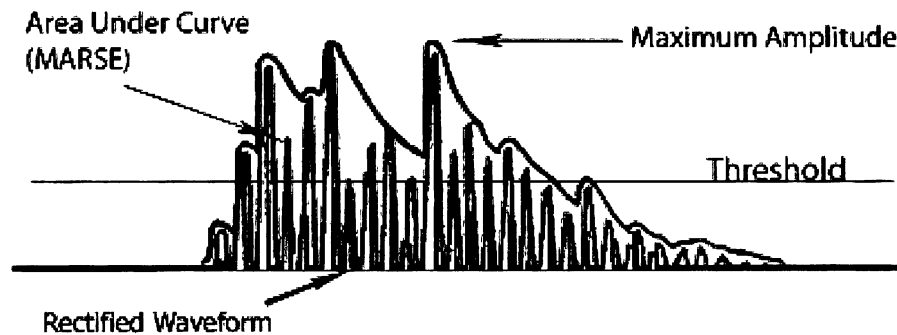

Figure 7: Mean area under the rectified signal envelope

The final parameter, which is used in this research but is generally not computed in other studies, is the mean frequency. Mean frequency can be expressed simply as the counts value divided by the duration. This average frequency value can be used in filtering of data during post-processing and is the basis for this research.

13

It should be noted that the circuitry that determines the parameters for the waveforms does not operate instantaneously; hence, if the parameters for a waveform cannot be computed before another waveform is encountered, the incoming waveform is sent to a buffer where it stays in its raw form until the system can determine its AE parameters All incoming waves are given a timestamp before entering the buffer (or if there are none in the queue, the analyzer circuitry) so that they are not out of sequence even if the amount of data in the buffer becomes quite large

After the AE parameters have been determined for each waveform, they are logged sequentially by the acoustic emission system   The logged waveforms and their parameters are stored in permanent storage, usually a hard drive of the acoustic emission system   From this location, they can be recalled, replayed, or otherwise reanalyzed at will during post-processing

A common problem with large acoustic emission data sets is multiple hit data   This type of event occurs when two or more waveforms arrive at a transducer within a short period of time   Here the second waveform arrives before the first waveform diminishes below the threshold and therefore causes the AE system to log both superimposed waveforms as one   This type of multiple hit waveform will have falsely identified quantification parameters that do not correlate with the rest of the single hit data   Multiple hit data must therefore be removed from the set   To better visualize multiple hit data, Figure 8 shows two waveforms that arrive in the AE system at nearly the same time
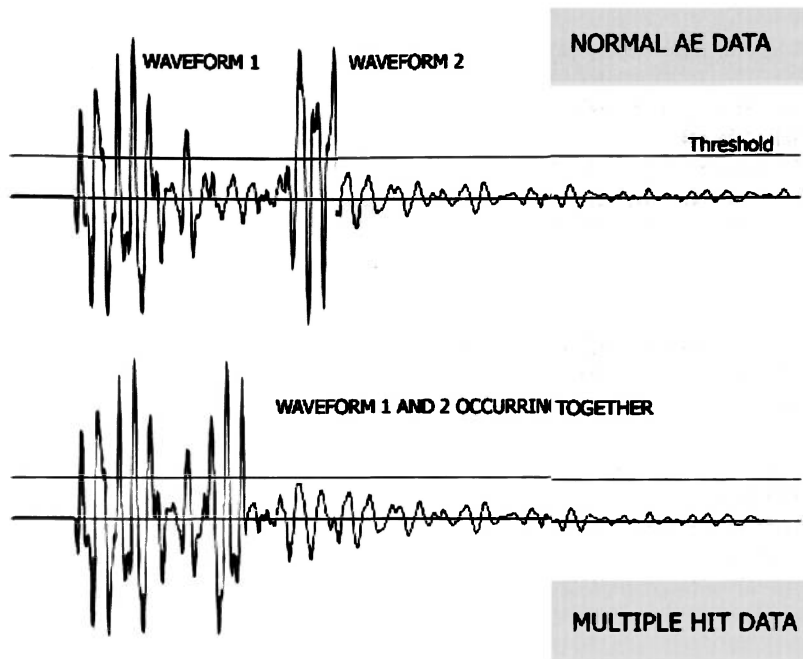
14

Figure 8: Regular AE vs. a multiple hit AE event

2.2 Artificial Neural Networks Overview

Artificial neural networks (ANNs) are a diverse set of robust mathematical tools used to classify data into clusters, recognize patterns, process signals, and do predictive modeling and forecasting. ANNs are patterned after the mammalian brain. This sort of arrangement is accomplished by interconnecting numerous artificial neurons called processing elements (PEs). A single PE has very little processing capability; however, when many are interconnected, a massively parallel system can be constructed that is capable of biological learning and problem solving. ANNs share many similarities to their biological counterparts, such as learning, memory, and judging new data based on previous experience.

ANNs and biological systems do differ, though: While the human brain is estimated to contain 100 billion neurons, artificial neural networks are currently much, much smaller. This is due to the fact that although the human brain is massively parallel, it operates at a maximum of 30 Hz whereas common modern computer processors approach 4 GHz. The computer's speed is over 100 million times as many calculations per second. This high speed processing capability has led to resurgence in the use of artificially intelligent systems where network sizes are not required to have anywhere near as many neurons as the human brain to have the same capabilities.

ANNs thrive at solving problems where there is no explicit solution or there are multiple interactions that are too complicated to separate and/or express with a known function.

16

The first network type utilized in this research, the backpropagation neural network, is used to generate predictions based on patterns it can find in the aforementioned variable interactions. To be more specific, this research's neural network analyzes the complex patterns in acoustic emission amplitude histograms and makes predictions of failure strength of the materials. The use of backpropagation ANNs in this research involves comparing the predicted failure strength to actual failure strength in an attempt to validate that the data fed into the neural network were sufficiently well-filtered to remove extraneous noise. Low error in these predictions (< 5%) usually signifies clean acoustic emission data.

The second type of ANN used in validating the quality of acoustic emission data in this research was the Kohonen self-organizing map. SOMs use a multi-dimensional matrix of network weights that respond to input values as well as other weights near the weight in question. This process has the overall effect of shifting like-valued weights together and reducing other weight values to zero. The end result is a matrix of weights that have well-defined groupings and that correspond closely to the major features in the data set. In the case of the materials-oriented testing in this research, the major features are the failure mechanisms in the data. If the acoustic emission data are clean, well-defined patterns in the weights associated with each failure mechanism will be evident.

2.3 Previous Research

Acoustic emission specialists have asserted for a long time that the greatest advantage of AE testing is also its greatest disadvantage This statement refers to the sensitivity of AE testing, which, while allowing for the detection of minute changes in a material, also allows for the introduction of a large amount of non-usable signals and noise into the data collected Filtering mechanisms of some kind are therefore necessary so that the relevant acoustic information can be sorted from the rest of the data.

Historically, filtering was done by controlling the frequency bandwidth that was sent to the AE analyzer system through appropriate selection of transducer and band pass filters placed on the amplified signal These rudimentary analog methods have been successful, but there are several difficulties associated with them First, the decisions for transducer and filter application must be made by a highly-trained individual These setup choices will vary from expert to expert and application to application Next, a trained individual must be present for analysis to look for recognizable patterns in the data amidst a variable level of noise Finally, in cases where a large amount of noise is generated, even trained individuals have a difficult time picking out failure modes and mechanisms of a material

In recent years, growth of AE usage in more complex (and noisy) materials systems has prompted researchers to look for new ways to extract a meaningful signal from the data A brief look at the three most widely used methods is appropriate here

Manual Filtering – Filtering AE data manually based on previous AE or material system knowledge is probably the most widely used technique. In some cases this method is done without considering the fact that the researcher is filtering the data at all. The manual filtering method usually involves removal of data points that have characteristics that lead the research to believe they are not characteristic of the type of signal expected. Some examples would be the removal of longer (or shorter) than expected durations, removal of long (and short) rise times, zero energy events, etc. These types of decisions are generally regarded as obvious to most users of AE and have been referred to as the "obvious" or "dummy" filtering done after AE data is taken. Some AE systems support setting parametric constraints, so that this type of filtering can be built into a test at the AE analyzer level.

Statistical Analysis Filtering – Knowing the ideal characteristics of AE data parameters in advance of gathering data has led some researchers to attempt to use statistical analysis and modeling techniques to aid in the removal of noise. For instance, an AE amplitude distribution usually takes the shape of a right-skewed beta distribution; researchers can curve-fit the data and remove outliers that would cause the data to have a poor fit. Another technique involves fitting a linear curve to a "duration vs. counts" plot, which should be linear, assuming a single prevalent failure mechanism. In this case, points that do not fall within a set number of standard deviations from the line can be considered outliers and removed from the data set.

Neural Network Filtering – The increase in computing power over the past decade has led to faster, more powerful neural networks Some neural networks are built to classify and group similar types of data together Specifically, Kohonen-type learning used in SOMs can usually separate AE sources into well-defined groupings of data points in n-dimensional space This allows for the removal of outliers that do not fit into the defined data groupings, as well as per-mechanism type analyses not possible with most manual or statistical techniques

Numerous researchers have worked with complex manual and neural network methods It has been found that in both cases, the filters and/or network must be changed and tailored to each type of test and material system It has also been shown that a large amount of human intelligence and time must be spent on these methods, thus leading to this attempt to develop a filtering technique that is more self-contained in its operation

3.0 METHOD

3.1 Filter Background

Several different aspects of the data and the way they were acquired led to the idea of filtering using the principal frequencies in a data set. The first aspect that led toward the ultimate use of frequency analysis in filtering was the relative ease with which the overall cleanliness of an acoustic emission data set could be determined from the duration vs counts plot. An ideal duration vs counts plot will display linear data sets corresponding to the most prominent frequencies found in the acoustic emission signals. It was surmised that if a valid approach of throwing away data that did not fall along these lines was available, it would leave only the acoustic emission data that were representative of the mechanisms at work within the material.

The second aspect that was considered has more to do with the acquisition method and equipment than the actual data. This consideration was with regard to the frequency response of the transducers used in the test. Acoustic emission transducers rely on piezoelectric materials to pick up the small vibrations in the specimen, and the piezoelectric material (which could be of varying size and geometry) is susceptible to different interference and harmonics that make it easier to acquire certain frequencies and more difficult to acquire others. This essentially means that the frequencies that are favored by the transducer and/or structure are going to carry the most information about what is actually going on inside the specimen. It is logical to assume that if there are principal frequencies in the data set, the transducer response is factored into those, and

this is therefore a positive aspect of the data gathering process. Transducers that have areas of greatly favored response, or areas that have peak response, such as resonant frequencies in a narrowband transducer, are never extremely high or low in the frequency response range of the transducer. This is a positive aspect, because most noise is expected to be in the low frequency (mechanical noise, multiple-hit data) and high frequency (EMI) ranges. This idea reinforces the fact that the areas of favored response will not acquire an overwhelming amount of noise, which would taint the entire acoustic emission data set.

COMPARISON OF SE150-M WITH SE1000-HI AND SE1000-H
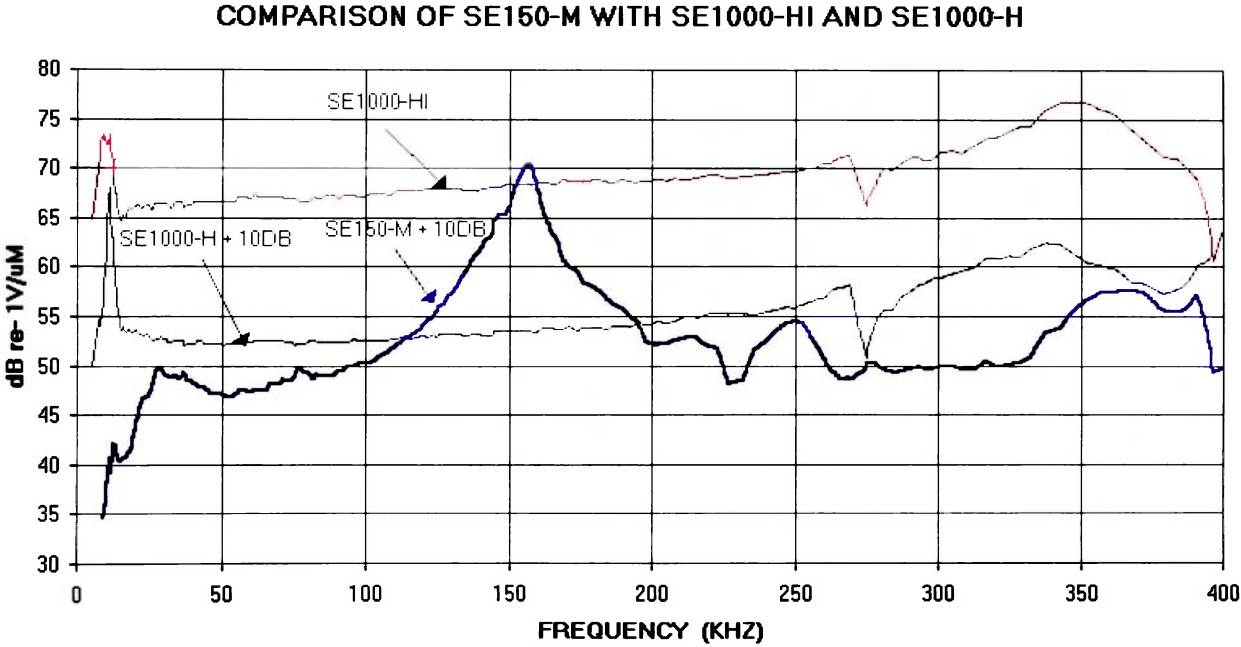
Figure 9: Sample transducer frequency response graphs

The blue transducer in the example frequency response curves of Figure 9 shows a common occurrence in the respect that 150kHz can become a favored frequency in the

spectrum, but otherwise these transducers have nearly equal response across all frequencies.

The final aspect considered deals with the mechanics of the material system and what actually produces acoustic emission. Because acoustic emission is produced by flaw growth activity, it is arguable that acoustic emission relies on the weak points in the material to produce emission indicative of the worst case in the specimen. This case is our primary interest, because the weak points (areas of high acoustic emission activity) are the points where failure is going to occur. Two conclusions can be ascertained from this information. First, the makeup of the material system does not matter: as long as a filter is applied that extracts the principal frequencies, it will be extracting the frequencies associated with the failure mechanisms because those will be most prominent. Second, specimen geometry does not matter: the weak points in the material may be in different locations and therefore different amounts of damping will occur on the way to the transducer, but as long as the signals reach the transducer they will be recognized as important frequencies (whatever they may be) because they occur more often than others.

The way the weak points in a specimen produce large amounts of acoustic emission also has another positive effect that was initially overlooked. Noise and multiple hit data generally come from many frequencies, which spreads it out all over the spectrum. The nature of some tests may cause this noise to be more pronounced at low (as in the COPV tests) or high frequencies, where it is entirely possible for a principal frequency to exist.

The fact that the weak points in the material produce so much activity at certain frequencies means that these frequencies, although surrounded by noise, are easily sorted out by the obvious peaks in the frequency response plot seen in Figure 10.

Frequency Distribution of Original Data
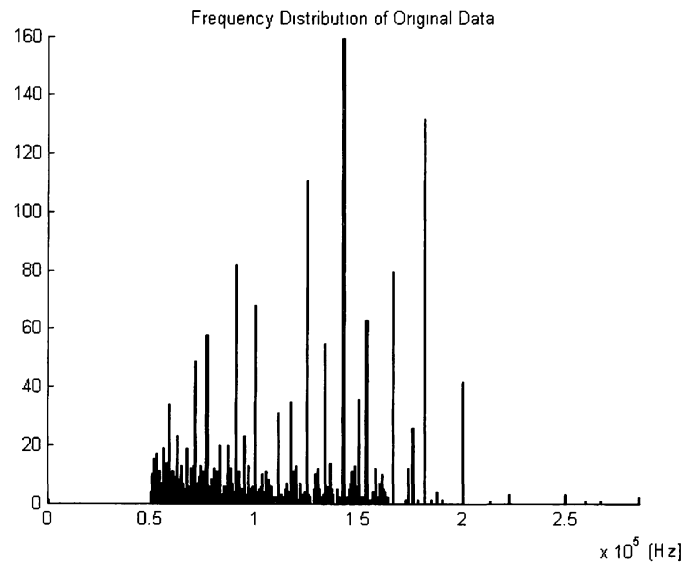
Figure 10: Sample frequency distribution

The primary frequencies in this example specimen are evidenced by peaks at 142kHz, 180kHz, and 125kHz. Note that there are peaks lower in the frequency range where a great deal of noise is expected, but the noise does not form peaks -- instead it is spread out in the 50kHz to 170kHz range and is unable to conceal the more prolific signals in the overall frequency distribution.

3.2 Utilization of the Filter

This section is structured to give a step-by-step representation of what the frequency-based filtering program does when it is run. This program was created in MATLAB, and the code can be found in Appendix A.

3.2.1 Step 1: Import Data File

The first step is to import the data into MATLAB. Modern acoustic emission systems output AE parameter data in tab-delimited column format in a text document. This can be imported directly into MATLAB with the data import function, as long as the data is oriented as follows:

Table 1: Data file architecture

Column	Parameter
1	Data point
2	User-defined parameter
3	Channel
4	Rise time
5	Counts
6	Energy
7	Duration
8	Amplitude

This data format is like what would be found in a text file such as the example shown in Figure 11.

point	press	ch	rise	counts	energy	dur	amp
3	0.02	6	13	12	15	115	69
4	0.02	5	17	119	230	1125	77
6	0.02	1	12	6	3	24	65
7	0.01	1	10	4	3	32	66
8	0.02	5	2	1	1	10	61
11	0.02	7	3	9	21	325	64
15	0.02	3	17	124	384	1054	83
18	0.02	5	2	5	15	235	64
21	0.02	1	4	11	10	90	69
27	0.02	1	20	4	6	72	62
34	0.01	7	6	14	27	375	67
44	0.01	6	7	2	4	65	62
45	0.02	1	15	1	1	17	60
46	0.02	4	2	3	5	72	63
48	0.02	2	4	8	8	64	70
49	0.02	5	25	13	15	69	69
52	0.01	5	3	6	27	340	62
54	0.02	5	2	3	2	21	65
60	0.02	1	10	2	2	14	64
61	0.01	2	1	2	1	16	62
66	0.02	5	12	11	15	166	64
67	0.02	5	13	5	8	76	66
69	0.02	4	11	10	9	95	66
70	0.01	4	13	10	10	79	67
73	0.02	1	3	2	1	15	63

Figure 11: Excerpt from sample data file

In the case of this example, the user-defined parameter was pressure, which was used in the COPV testing. The test took data from seven AE transducers, and some preliminary data reduction has already been done to lower the number of data points, as evidenced by the data points missing from the first column, which were removed by a computer to reduce the data file size.

3.2.2 Step 2: Set Filter Program Parameters

There are two user-defined parameters that need to be set in the program before it can be run. The variable "frequencyrange" will set the size of the "sorting bins" used in the frequency spectrum analysis. This variable has a default setting of 700Hz for composite

26

specimens, which means that the program will consider that any two frequencies that are different by less than 700Hz can be grouped together. The default values for composite (700Hz) and metallic (500Hz) specimens were determined empirically and work well for the materials and specimens used in this research. The variable "freqstofind" (frequencies to find) is an indicator of how many frequencies are to be determined by the program as principal frequencies. This variable has default values of 10 for composite specimens and 5 for metallic specimens. Generally, there are 5 or less failure mechanisms in the specimens analyzed, but there are other areas of high activity that were determined to carry useful information. It should be noted that the code currently only analyzes Channel 1, which was determined to carry the most information. The Channel 1 transducer on the COPV specimens was mounted on the polar boss of the specimens, and the Channel 1 transducer for the metal specimens was mounted on the upper grip of the tensile testing machine. Future versions of this program will determine the most active channel automatically.

3.2.3 Steps 3-7: Running the Filter Program

(Step 3: Preliminary Calculations)

Because frequency is not one of the commonly output acoustic emission parameters, it is necessary to first loop through all of the data points and calculate the frequency associated with each hit. The first part of the program divides the information from column 5 (counts) by that from column 7 (duration) of the imported data and thereby calculates the mean frequency of the waveform. Column 5 is counts expressed as an

27

integer, and column 7 is the duration in microseconds, so column 7 must be converted from microseconds to seconds in order to have the frequency in the base unit of Hertz (cycles per second). Note that this calculated frequency is identified as the mean frequency, due to the fact that it may change throughout the duration of the waveform. For the sake of simplicity, the mean frequency will be hereafter identified only as "frequency" in this research.

(Step 4: Initial Filter)

The first actual filtering of the data takes place in a part of the program that applies some techniques based on acoustic emission-driven logic. This initial filter steps through all of the data points and removes data in the auditory and near-auditory ranges (0 to 50kHz) as well as any data above 2MHz, which is generally known to be outside the normal range of mechanisms detectable by acoustic emission. During the primary check for falling within the acceptable frequency range, any zero energy events, which contain no usable data, are also removed.

(Step 5: Principal Frequency Determination)

During the frequency determination phase, the program orders all of the frequencies encountered during the initial filtering from largest to smallest frequency of occurrence. After this is complete, the program picks out the number of frequencies to isolate as determined by assignment of the "freqstofind" variable and stores those as the principal frequencies.

(Step 6: Data Extraction)

Utilizing only the data points that correlate with the principal frequencies would oftentimes extract a data set that does not have sufficient data points from which to draw any statistically significant conclusions. Therefore, the program uses the principal frequencies as a starting point and extracts data from the principal frequency as well as data within a range of ±2% around the principal frequency. The 2% value is determined by dividing the total frequency range observed in the data set by 50. After the extraction procedure has been performed for each of the selected frequencies, the data is prepared as a new set for creating plots, as well as potentially being output to a file or another program.

(Step 7: Generation of Visual Analysis Plots)

The plots used in the visual analysis are generated at this final phase in the filtering process. After plotting the principal frequencies, a duration vs. counts plot of the original data set is plotted, with an overlaying plot in a different color of the new data set so that an immediate comparison can be made between the original and the edited data. An amplitude histogram is generated of the original data set with an overlay of the edited data set, but because the edited data set is often orders of magnitude smaller and more difficult to view, a separate plot of the edited data is generated for visual reference. Similarity in shapes of the original and filtered data plots indicate that no vital information was lost in the filtering process. Sample plots are shown in Figure 12.
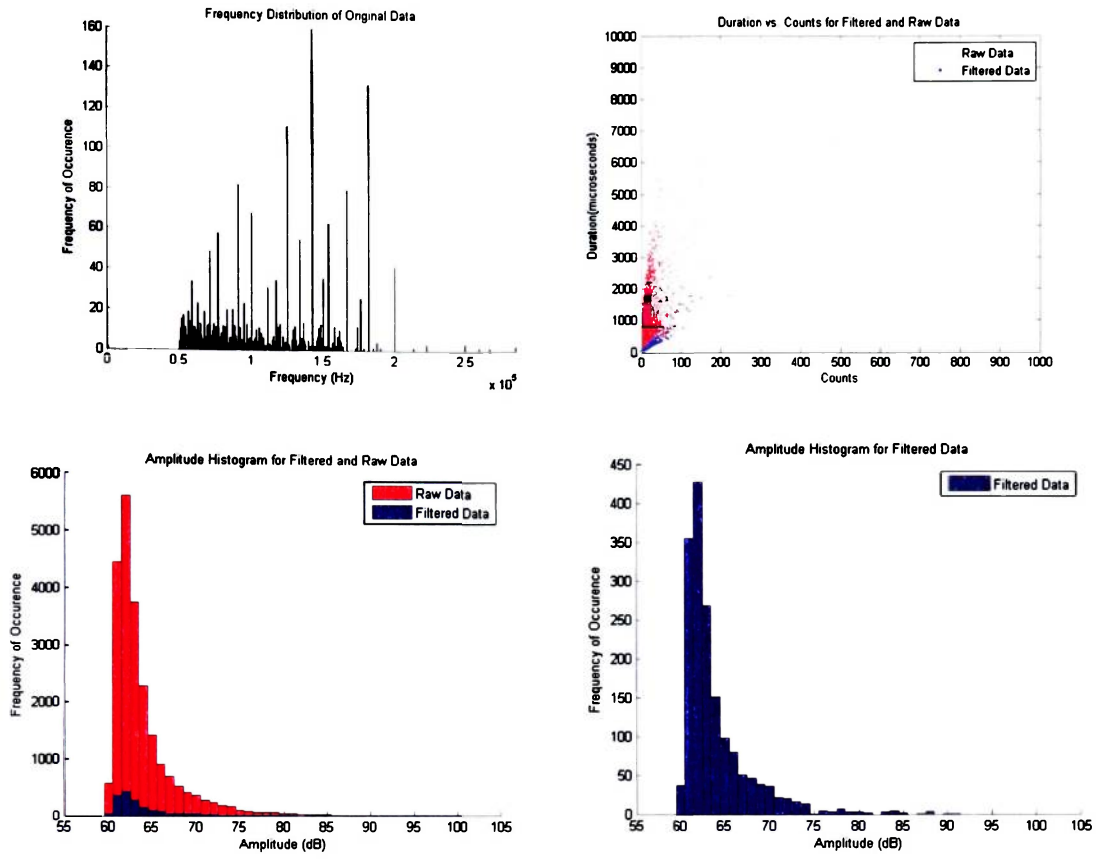
Figure 12: Sample filter program output

4.0 RESULTS

4.1 Visual Data Set Comparison

4.1.1 Composite Specimens

In the composite specimens, the visual data comparison results were clear. The unfiltered data were very difficult to match to any of the expected acoustic emission plot forms, but the filtered data seemed to match the theoretical cases very well.

The duration vs. counts plot for COPV-009 is shown in Figure 13. It is readily apparent that the data shown in red, which is the original data, is certainly not linear and does not have any separated linear segments that are discernible.
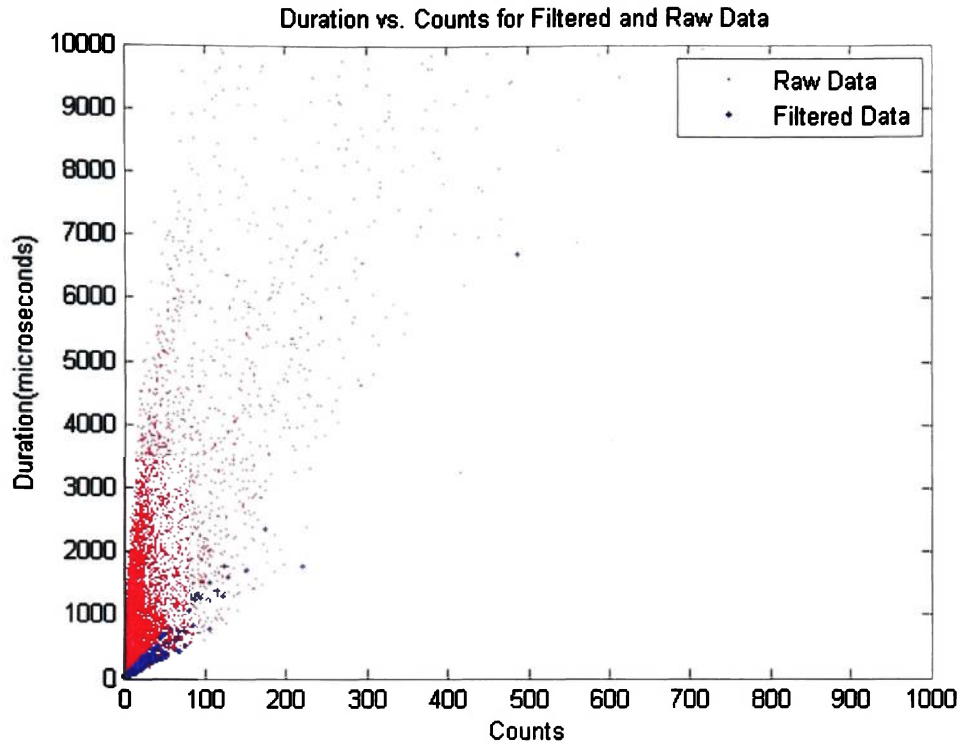
Figure 13: Composite specimen duration vs. counts output

It is also apparent that the filter has concentrated on 4 or 5 failure mechanisms while ignoring the large amount of low-frequency noise (the large red mass at the bottom-left of the plot) in this specimen. Zooming in on the origin of the plot confirms that the filtered data is from 5 mechanisms and that those points representing those mechanisms form five distinct straight lines.
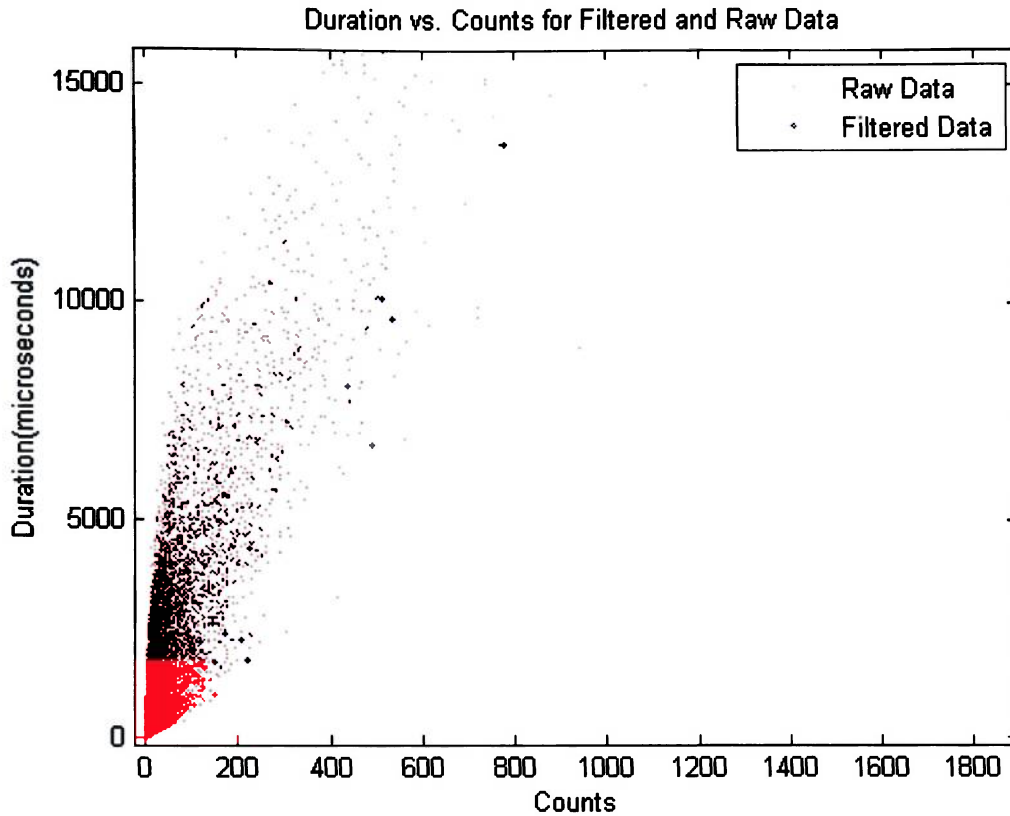
Figure 14: Composite specimen duration vs. counts output (zoomed to origin)

Analysis of the other plots immediately generated by the filter program will further confirm that the composite specimen data sets have been cleaned and appear to behave like idealized acoustic emission data. The plots shown in Figure 15 are first, an amplitude distribution of the old and new data sets on top of each other, and second, the new data set only.
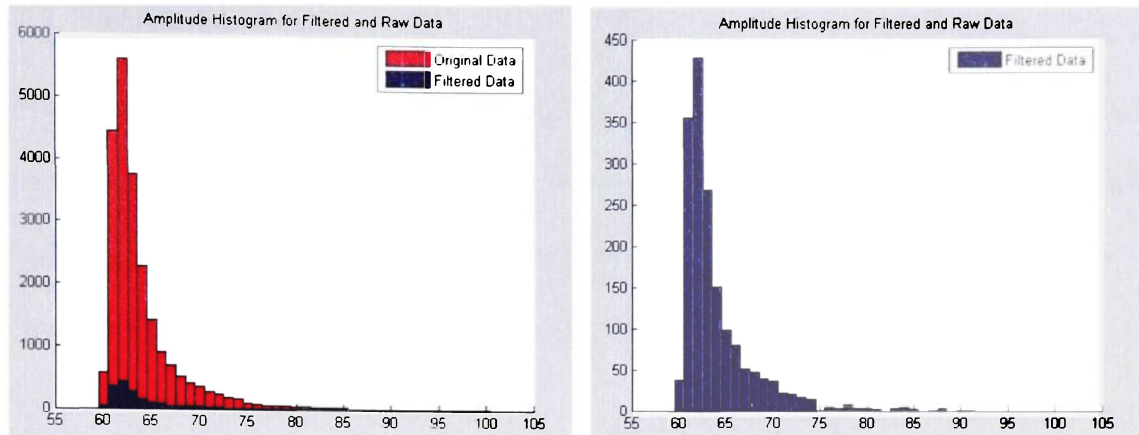
Figure 15: COPV unfiltered (left) and filtered (right) amplitude histograms

It can be seen from these plots that the original data set has little or no definition in the histogram that would indicate anything other than one large failure mechanism. This appears to be a consequence of the overwhelming number of hits in the first data set. The second data set, seen by itself on the right, has more definition in the main part of the distribution, showing one large and one less dominant failure mechanism between 60dB and 75dB. What it has that the original data set lacks completely, however, is the clear definition of the additional three failure mechanisms seen between 75dB and 100dB. It is also important to note the scale -- the filter has reduced the AE occurrences at 62dB from over 5500 hits to a more manageable 425. This leads to a data set that is more easily broken down into its constituent failure mechanisms. Thus, both human and computer based analysis of the reduced data set will be accomplished much faster and more accurately than with the original data set.

4.1.2 Metallic Specimens

Because the filter is theoretically based on principles that should not change with specimen geometry or makeup, the next important thing to do was to look for similar results in the metallic specimens. Shown in Figure 16 is the comparative duration vs. counts plot for Specimen 003 tested by Ibekwe in 2004.
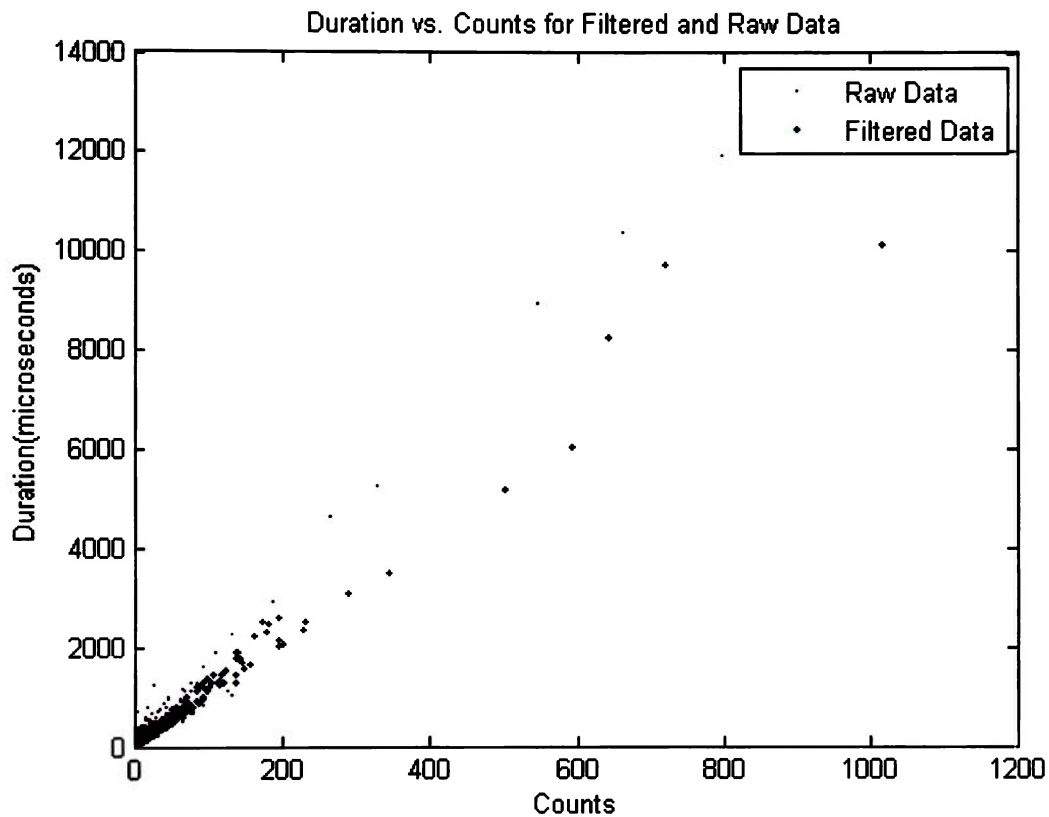
Figure 16: 7075 specimen duration vs. counts plot

As expected, the acoustic emission data from the metallic specimen does not look like the activity seen in the composite. However, the important effects of the filtering can

certainly be seen. There has been a significant reduction in data, and the highlighted blue region shows two very clearly defined linear failure mechanisms.
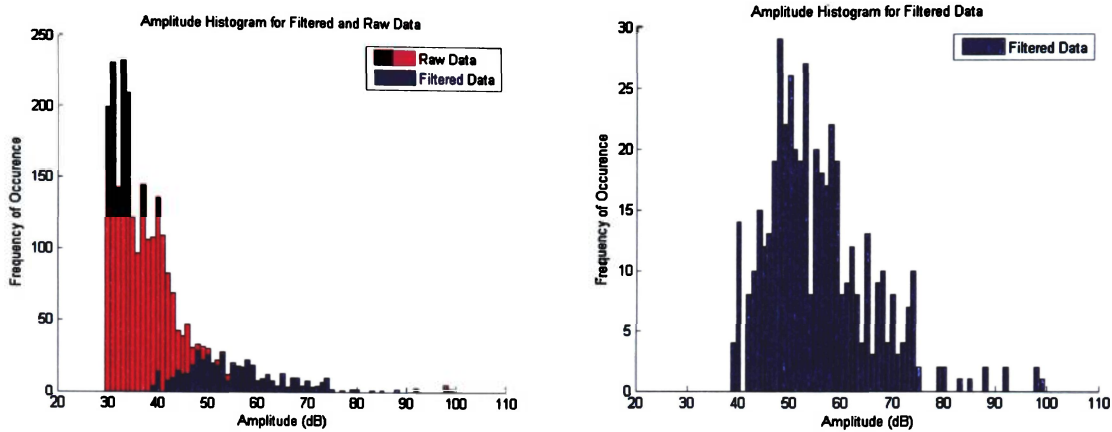
Figure 17: 7075 specimen unfiltered (left) and filtered (right) amplitude histograms

The amplitude histogram also shows very positive results. The original amplitude histogram was defined heavily by activity between 30dB and 40dB, but the filtering technique has revealed that all of that activity was some type of noise. The filtered histogram now shows two definite failure mechanisms at low amplitude levels, one of which is dominant. This filtered histogram has also revealed what could be a third, weak mechanism at slightly higher amplitude levels. The filter, in this case, has shown that the raw data set had valuable information about the failure of the specimen embedded in it that was not perceptible in the original data plot.

4.2 Neural Network Data Set Comparison

4.2.1 Backpropagation Neural Network Analysis

Due to the fact that the accuracy of backpropagation neural network predictions have been previously shown to directly relate to the relative amount of noise in the acoustic emission data, these results can be used as a gauge for the effectiveness of experimental filtering techniques. This research will analyze the burst pressure predictions generated by a network built in Neuralware's Neuralworks Professional II/Plus software using both the original and filtered data sets. As per the usual in the case of designing neural networks, the primary measure of success will be the quantity of the maximum error in the testing sets.

The first backpropagation neural network was run on data sets that had only basic manual filtering applied. Exceedingly low and high durations were removed, and short rise times were removed in an attempt to eliminate any electromagnetic interference. The results are shown in Table 2.

Table 2: Manually filtered composite data neural network prediction results

S/N	Damage	Purpose	Burst Pressure [psig]	Predicted Burst Pressure [psig]	Error
2	Impact	Train	1880	1892.014	0.64%
3	Impact	Train	2004	2076.125	3.60%
9	None	Test	2544	2138.136	-15.95%
10	None	Test	2460	2238.825	-8.99%
13	None	Train	2874	2888.485	0.50%
14	None	Train	2390	2429.957	1.67%
18	Lacerated	Train	2864	2845.314	-0.65%
20	Impact	Test	1967	1865.053	-5.18%
25	Lacerated	Test	2393	2052.189	-14.24%
26	Lacerated	Train	2675	2654.171	-0.78%

Although the manual filtration had been applied, the neural network was still producing a maximum error in excess of 15% in the testing set. The filtered data was then used to train and test a network of identical architecture, and the results improved considerably, as seen in Table 3.

Table 3: Results for network after automated filter

S/N	Damage	Purpose	Burst Pressure [psig]	Predicted Burst Pressure [psig]	Error
2	Impact	Train/Test	1880	1885.207	-0.276%
3	Impact	Train/Test	2004	1981.855	1.105%
9	None	Test	2544	2576.606	-1.281%
10	None	Test	2460	2345.234	4.655%
13	None	Train/Test	2874	2871.966	0.070%
14	None	Train/Test	2390	2397.802	-0.346%
18	Lacerated	Train/Test	2864	2859.827	0.145%
20	Impact	Test	1967	2057.608	-4.606%
25	Lacerated	Test	2393	2398.023	-0.209%
26	Lacerated	Train/Test	2675	2605.398	2.601%

In this case the network was able to predict within the more restrictive ±5% error requirement. The maximum error in the test set was 4.655%. Although lower values have been obtained with trial-and-error filtering on a specimen-by-specimen basis in the past, it is important to note that this result was obtained from an automated filtering technique without human intervention.

4.2.2 Self-Organizing Map Analysis

Previous research has shown that another type of neural network, the Kohonen self-organizing map, also has performance that depends greatly on the quality of the input data. The SOMs used in this research should theoretically shift their weight values into recognizable patterns that reflect the failure mechanisms resident within a given material. First, a SOM with a matrix of 20x20 weights was used on the composite specimens, as seen in Figure 18.
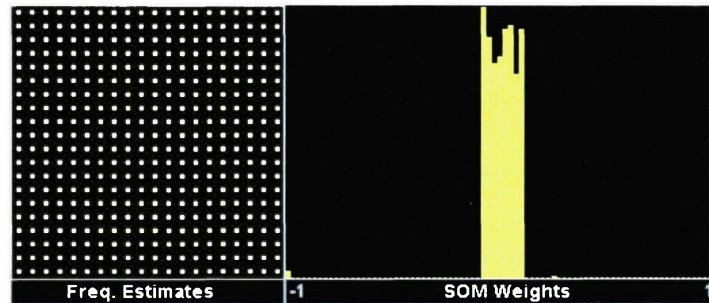
Figure 18: Frequency estimates and values of SOM weights

The frequency estimates (of the weight values) and the distribution of the weight values themselves show that no shifting of the weights from their randomly initialized values has taken place. In this case, the network was unable to do its job because the data were too noisy.
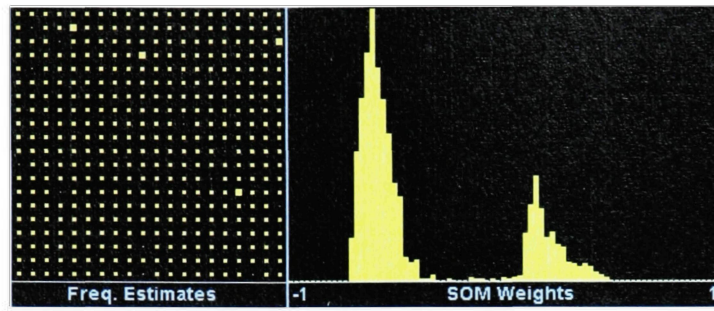
Figure 19: Frequency estimates and SOM weight values- filtered data

Figure 19 shows the results of the same SOM applied to the filtered data. Here the weights have shifted to show two major failure mechanisms, with what appears to be two smaller mechanisms as well. Because four mechanisms were expected, another SOM with a weight matrix of 1x5 (having a total of 5 possible outcomes) was run on the filtered data seen in Figure 20.
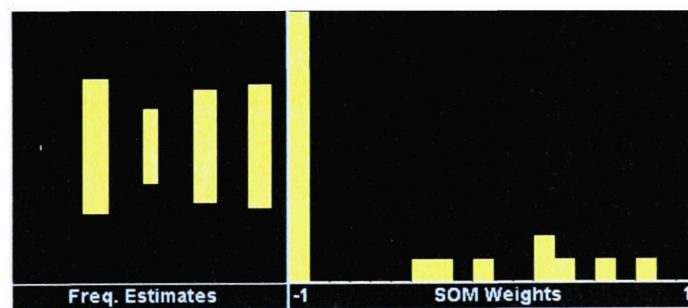
Figure 20: Frequency estimates and SOM weights for 1x5 size

Figure 20 shows a weight distribution that makes it hard to determine the number of mechanisms, but the frequency estimates reveal that there are four large weight groupings

and one extremely small one. This can be attributed to the main four failure mechanisms in the composite and a small amount of remaining noise.

The same 20x20 SOM that was initially applied to the COPV data was then applied to the metallic specimen data. Results of using raw data from Specimen 001 are shown in Figure 21.
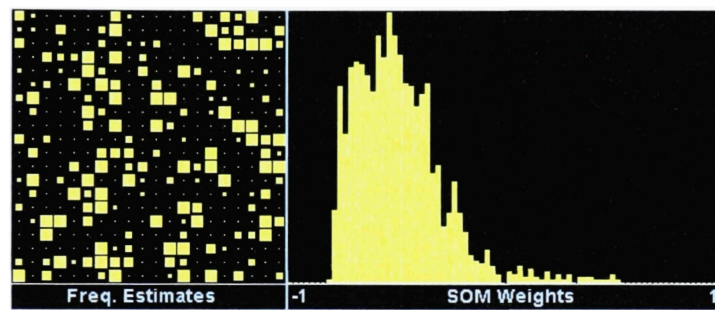
Figure 21: Frequency estimates and SOM weights for 7075-001, unfiltered

In this instance, the network was able to change the weights for the unfiltered data, but the grouping that occurred seems to suggest one large failure mechanism where two are expected. It was hypothesized that noise was obscuring the true nature of the data and therefore the filtered version of the data was run in the same network. These results are shown in Figure 22.
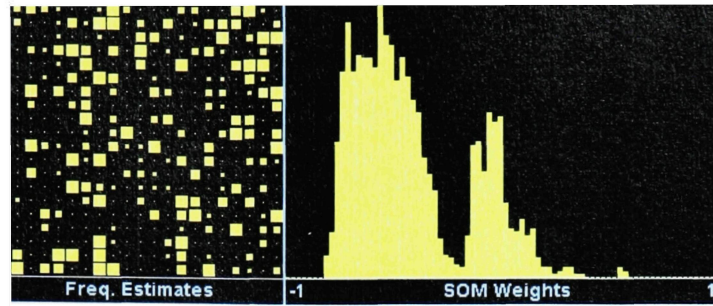
Figure 22: Frequency estimates and SOM weights for 7075-001, filtered

In Figure 22 it is easy to observe the effects of filtering on the data -- the network was able to reveal two obvious failure mechanism induced groupings in the weights, as expected, plus what appear to be two less frequently occurring mechanisms.

4.3 Additional Data Set Comparisons

The previous section discussed a number of validation methods, citing COPV Specimens 002 and 009 as well as metallic Specimens 001 and 003 for example purposes (all COPV specimens were used in the backpropagation neural network comparison).

This section will offer additional side-by-side data set comparisons for each specimen type. In the interest of keeping this section to a reasonable length, two data sets from each type were selected at random. All of the remaining data sets are contained in Appendices B and C.

4.3.1 COPV Specimen 003

COPV Specimen 003 exhibited several strong frequency components near 75kHz as well as the expected strong components between 125kHz and 150kHz, the favored range of the transducer.
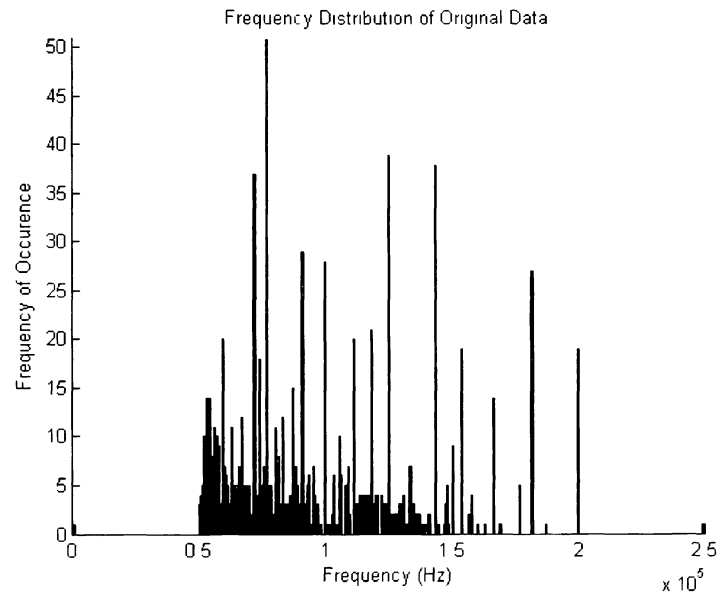
Figure 23: Frequency distribution of COPV-003 data
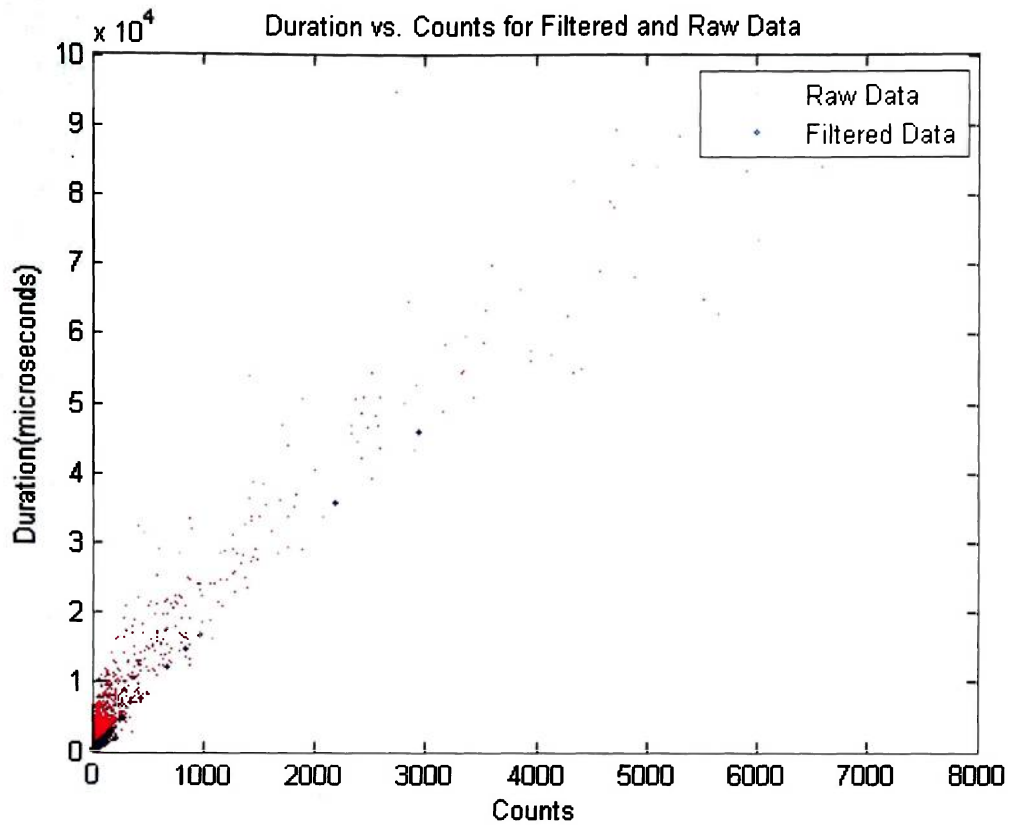
Figure 24: Duration vs. counts for COPV-003

The duration vs. counts plot (Figure 24) as well as the comparative amplitude histogram (Figure 25) shows significant reduction in data due to the filtering. The original amplitude histogram in Figure 25 appears to have a single large failure mode.
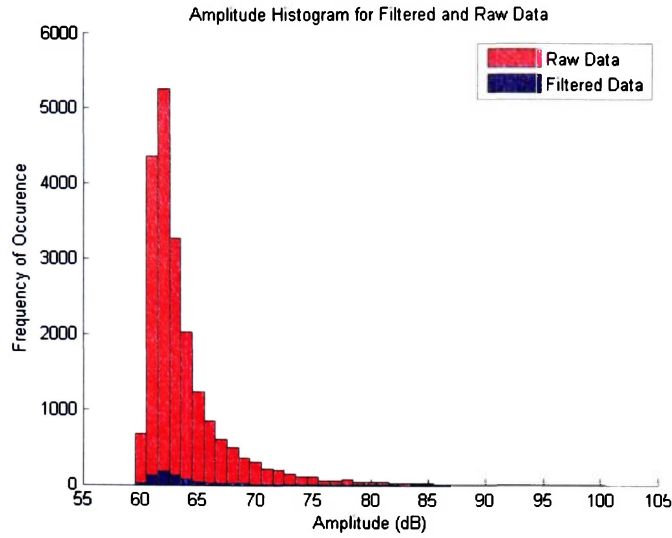
Figure 25: Amplitude histograms for filtered and raw data, COPV-003
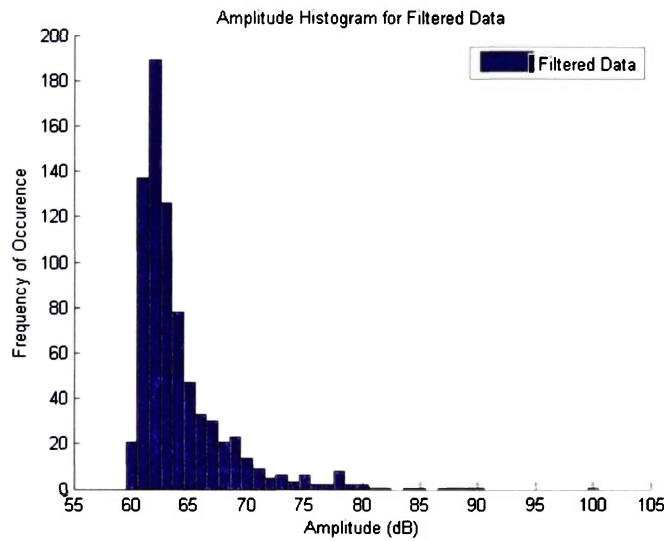
Figure 26: Filtered amplitude histogram for COPV-003

Figure 26 shows several well-defined failure mechanisms, including a large mechanism at 62dB and smaller mechanisms with peaks near 69dB, 78dB, 85dB and 90dB.

46

Side-by-side comparison of the unfiltered and filtered 20x20 SOM data for COPV-003 in Figure 27 shows an obvious improvement in data quality and structure for the filtered data. The unfiltered data would not shift the SOM weights from their initial values, while the filtered data shows two large failure mechanisms and several smaller ones.
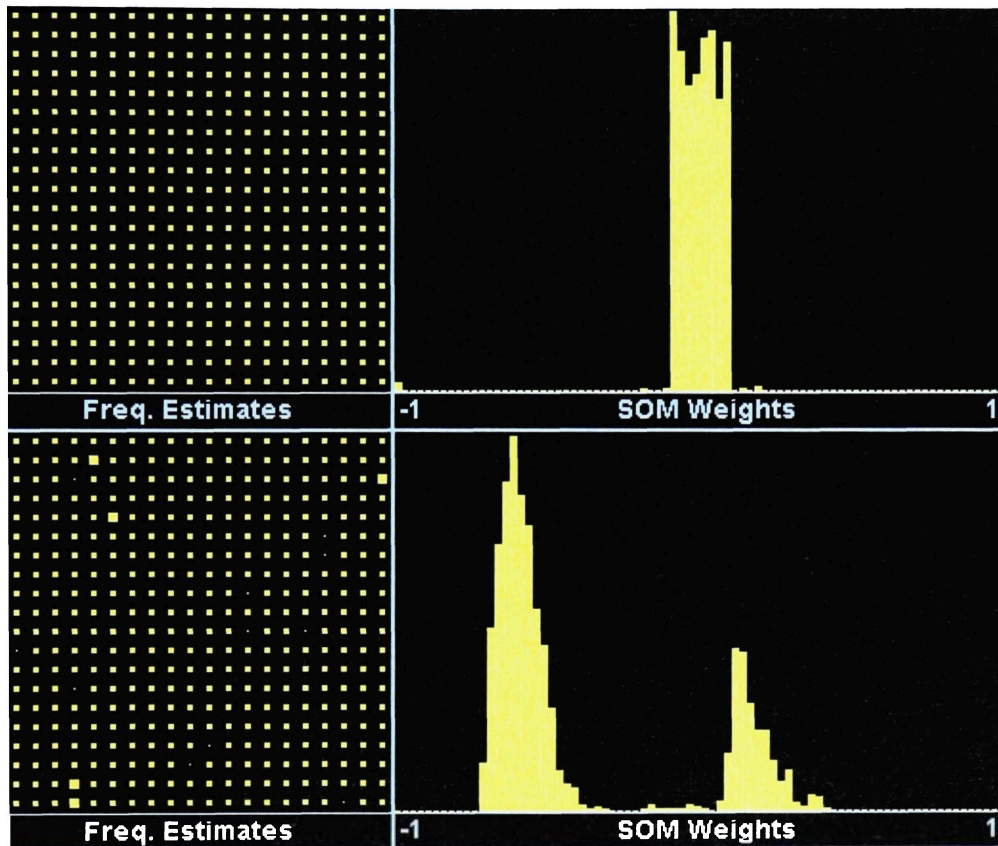
Figure 27: Unfiltered (top) and filtered (bottom) COPV-003 data using a 20x20 SOM

4.3.2 COPV Specimen 013

COPV Specimen 013 showed several frequency components in the range 125kHz-150kHz and appears to have approximately even spacing of the strong components in both the higher and lower frequency ranges.
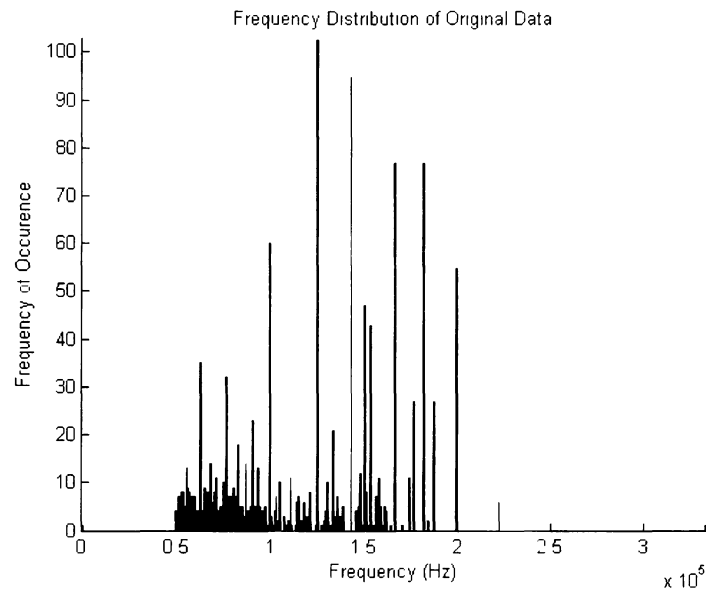
Figure 28: Frequency distribution of COPV-013 data
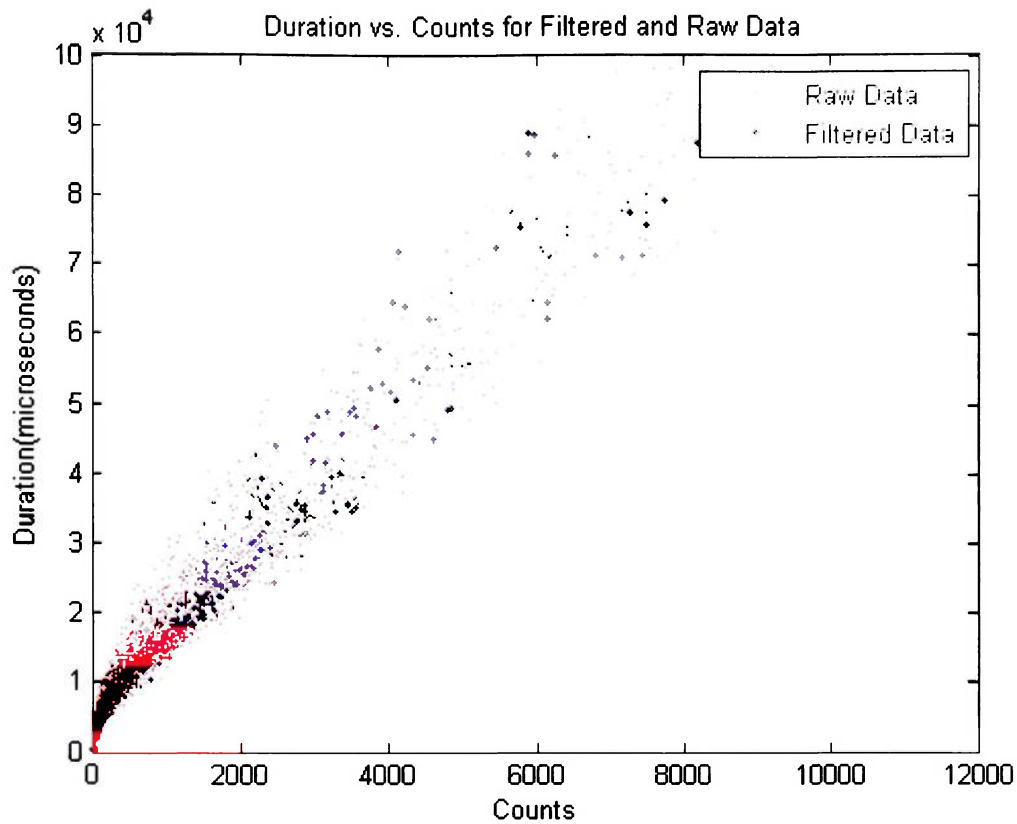
Figure 29: Duration vs. counts for COPV-013

The duration vs. counts plot seen in Figure 29 shows that the filtering technique has

isolated a number of frequencies from a great deal of low level noise. The comparative

amplitude histogram, (Figure 30) once again shows a significant reduction in data. The

raw amplitude histogram in Figure 30, like most of the COPV specimens, displays what

seems to be a single large failure mode.
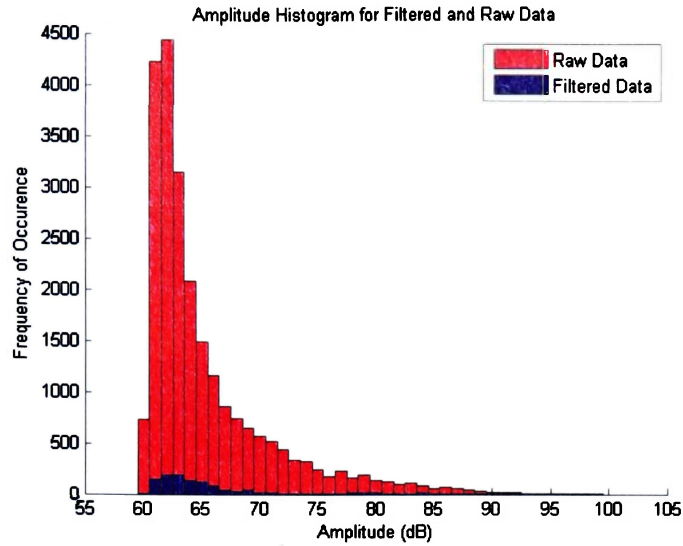
Figure 30: Amplitude histograms for filtered and raw data, COPV-013
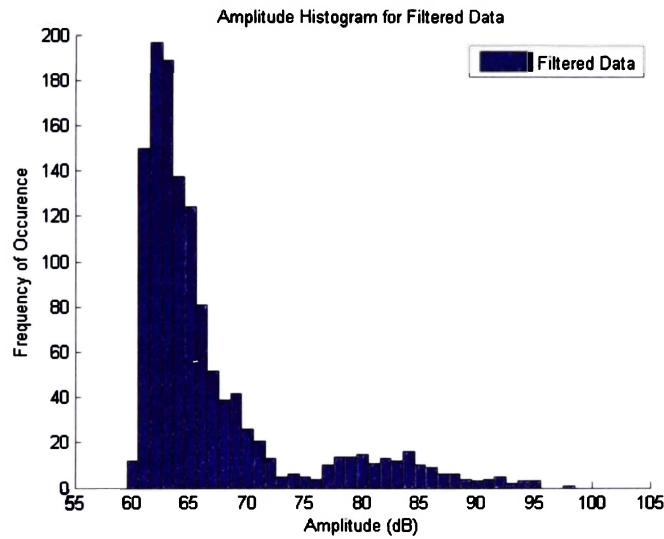
Figure 31: Final amplitude histogram for COPV-013

Figure 31 once again reveals that the specimen had several failure modes instead of the one large mode perceived without filtering. COPV-013 appears to exhibit separate modes at 62dB, 69dB, 75dB, 80-85dB, and 92dB.

50

Side-by-side comparison of the 20x20 SOM data for COPV-013 shows another major improvement in data structure. The SOM was again unable to shift the initial weights unless the filtering technique had been applied. The filtered SOM clearly shows weights shifted to 2 major groupings and 3 minor ones.
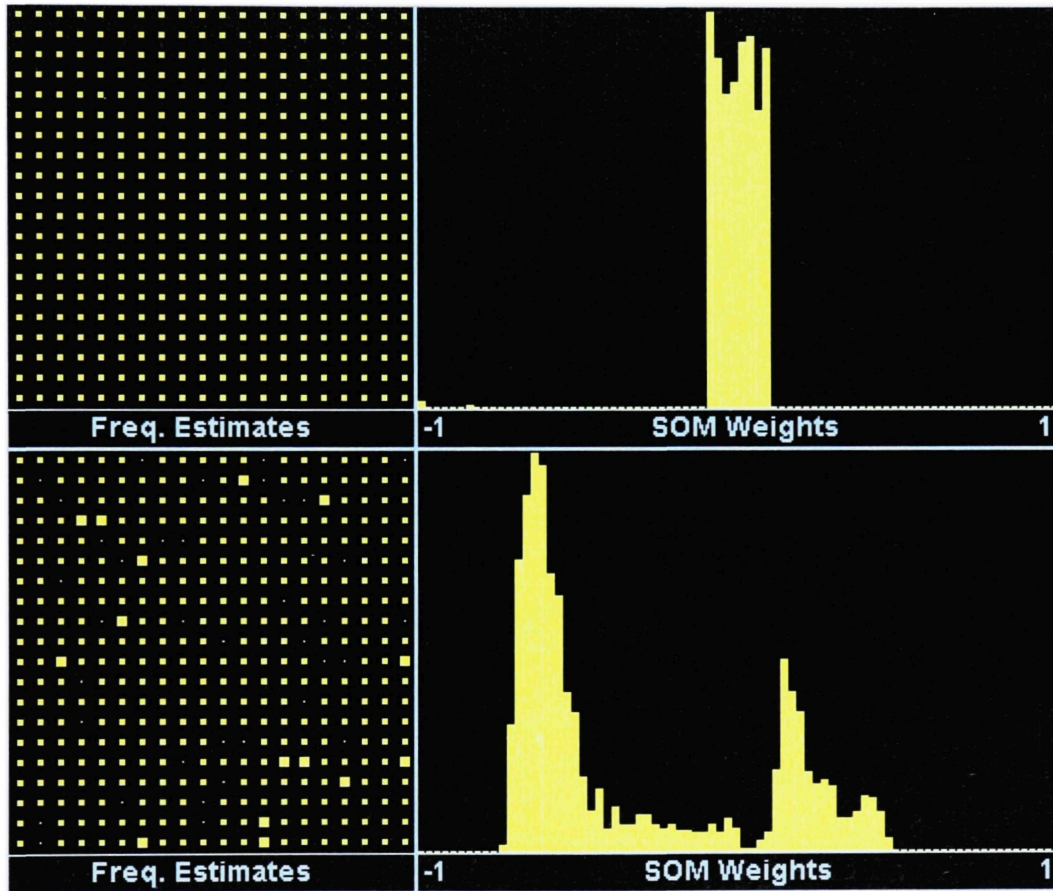
Figure 32: Unfiltered (top) and filtered (bottom) COPV-013 20x20 SOM

4.3.3 7075 Specimen 007

From Figure 33, 7075 Specimen 007 appeared to have approximately even spacing of strong components in mostly lower frequency ranges.
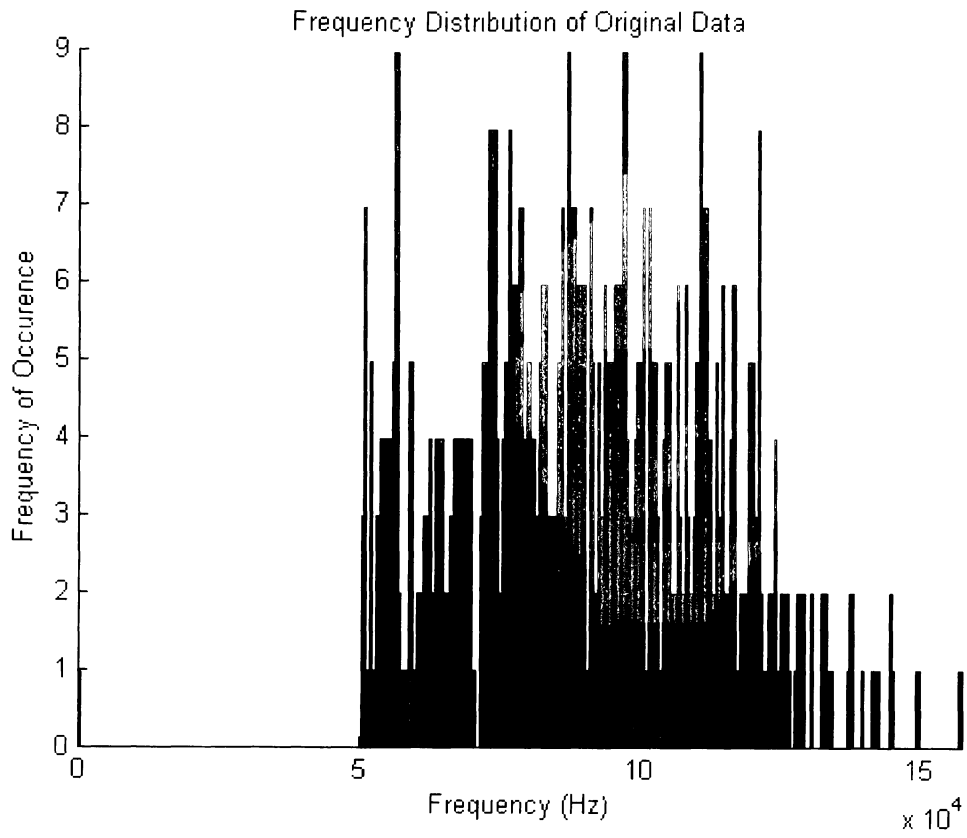
Figure 33: Frequency distribution of 7075-007 data
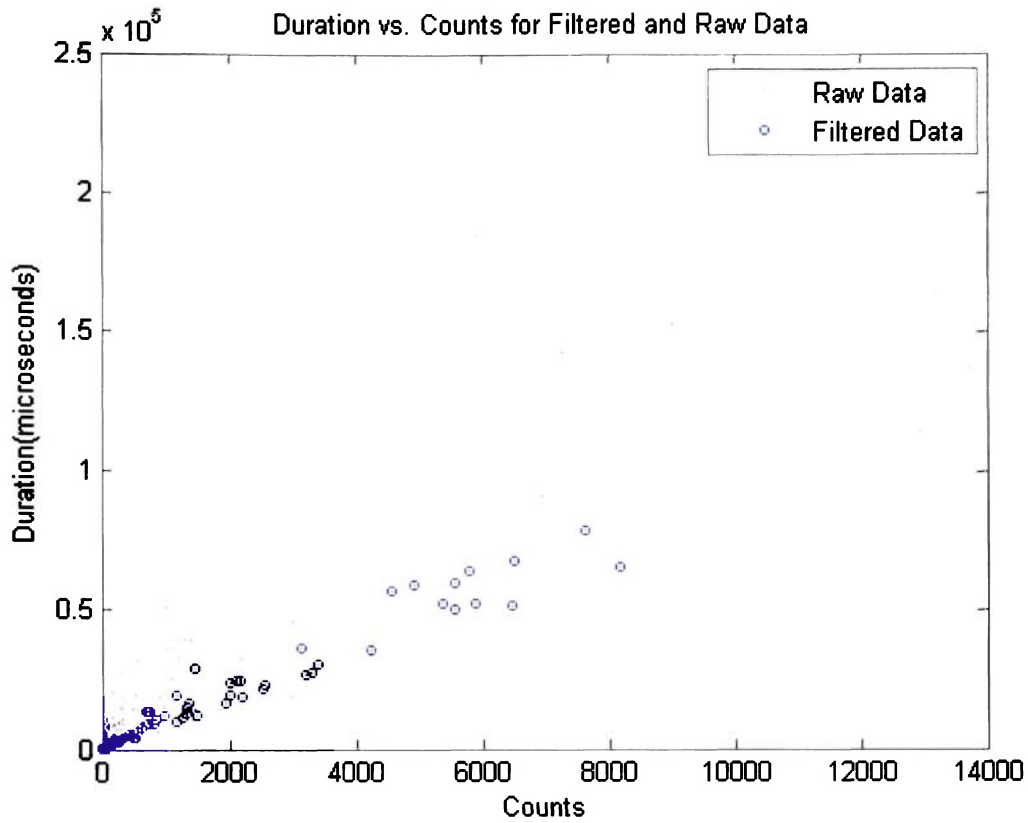
Figure 34: Duration vs. counts for 7075-007

The duration vs. counts plot seen in Figure 34 shows that the filter has eliminated approximately two-thirds of the lower frequency range data. Figure 35 once again shows this significant reduction in data as well as the range within which it occurs. The raw amplitude histogram in Figure 35, displays a single large mode that was mostly comprised of this low level, low frequency range noise.
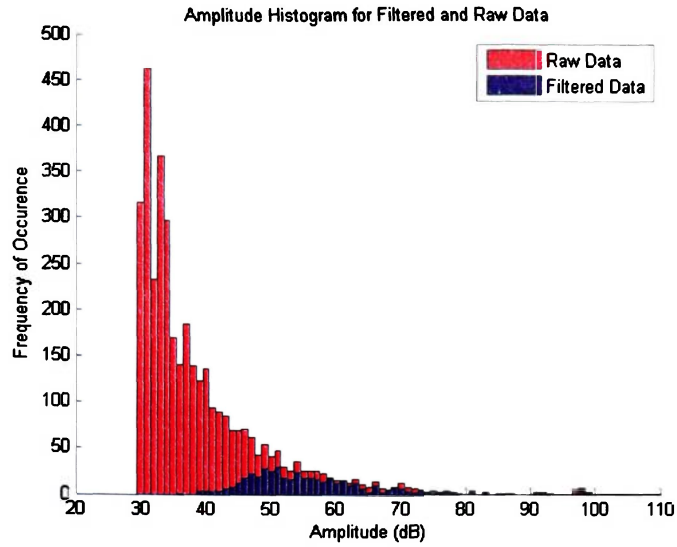
Figure 35: Amplitude histograms for filtered and raw data, 7075-007
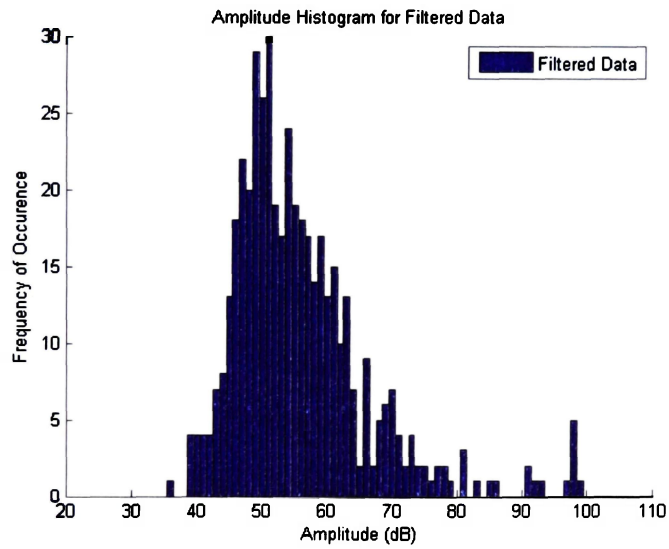
Figure 36: Final amplitude histogram for 7075-007

Figure 36 shows that Specimen 007 had several failure modes obscured by noise, including two major modes at 50dB and 70dB, as well as some significant occurrences at

higher amplitudes. Side-by-side comparison of the 20x20 SOM data for 7075-007 shows significant improvement in data structure. The SOM was unable to shift the initial weights before the filtering technique had been applied. The filtered SOM clearly shows weights shifted to 2 major groupings.
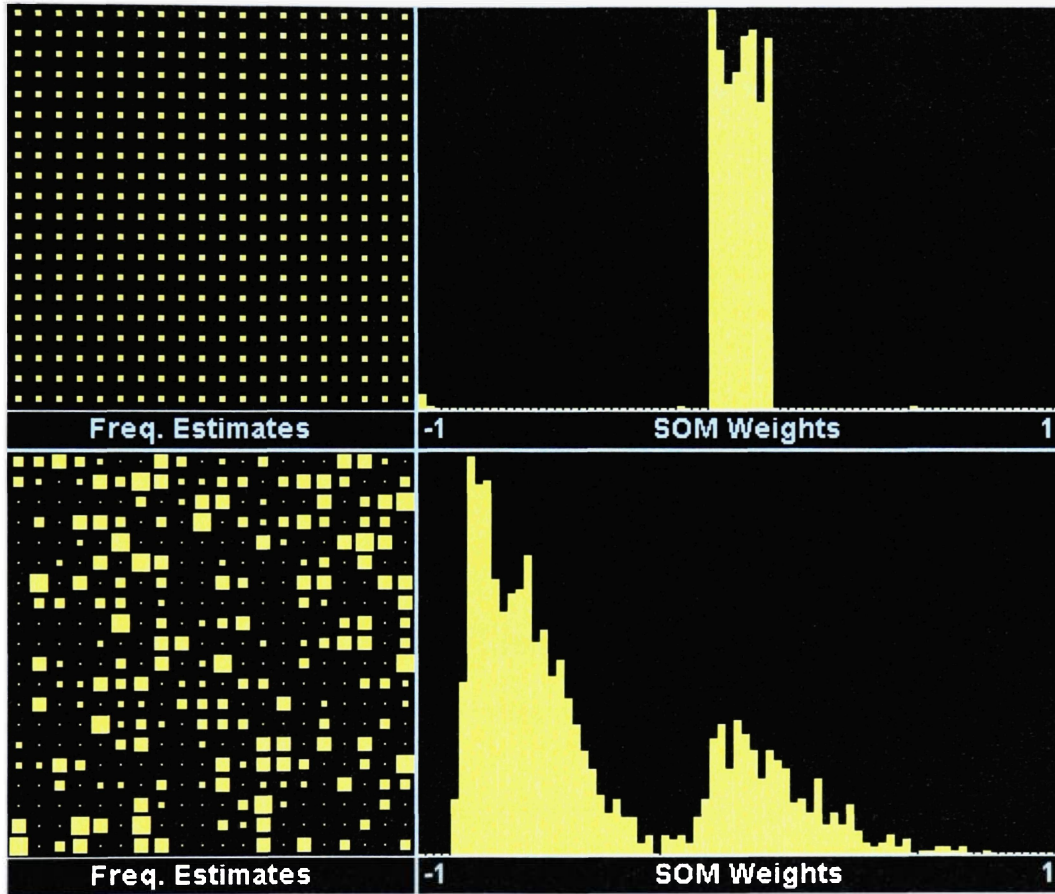
Figure 37: Unfiltered (top) and filtered (bottom) 7075-007 20x20 SOM

4.3.4 7075 Specimen 011

7075 Specimen 011 had most of its data, as well as its most prominent frequencies, at low

frequency ranges between 60 and 120 kHz, as seen in Figure 38.
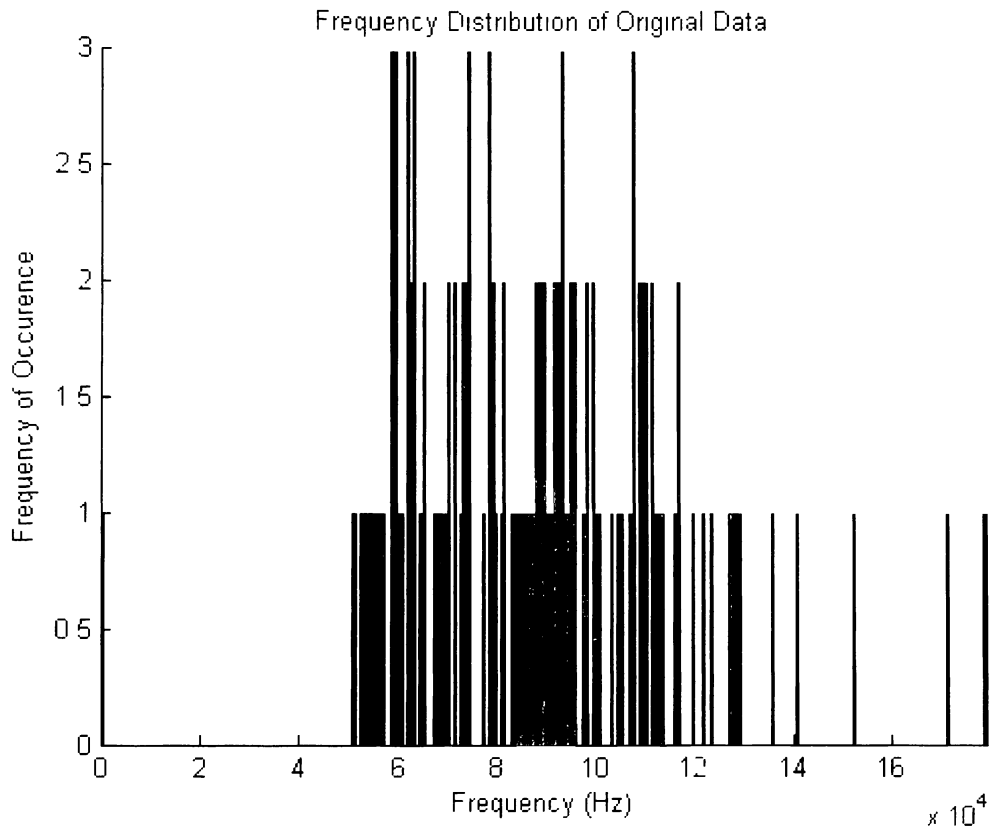
Figure 38: Frequency distribution of 7075-011 data
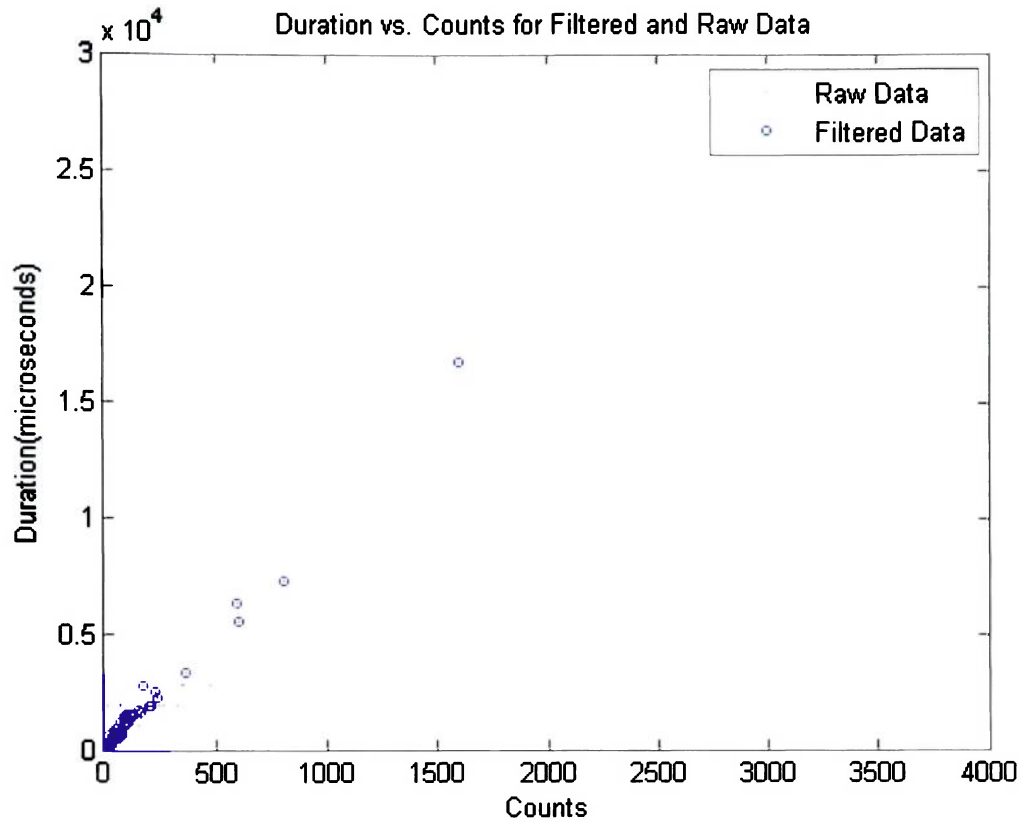
Figure 39: Duration vs. counts for 7075-011

The duration vs. counts plot seen in Figure 39 shows that the filter has deemed approximately half of the data above and below the average frequency to be noise. Although a large reduction was not accomplished in this case, the following plots will show that significant improvement had been made to the structure. Figure 40 shows a very significant amount of noise below 45dB in the unfiltered data.
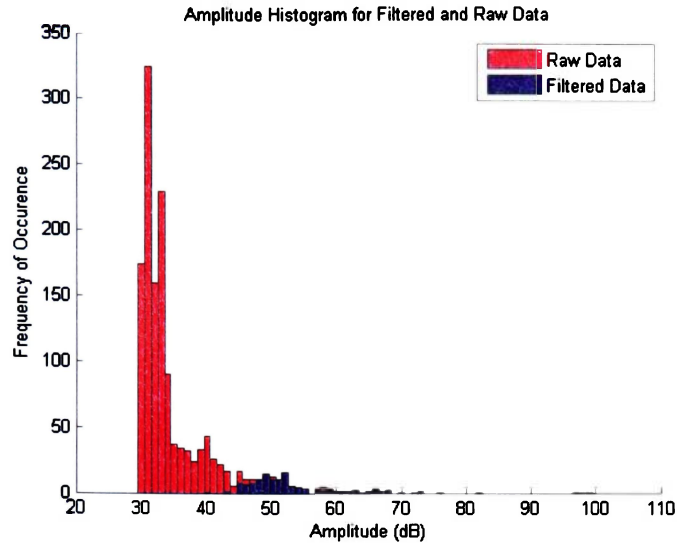
Figure 40: Amplitude histograms for filtered and raw data, 7075-011
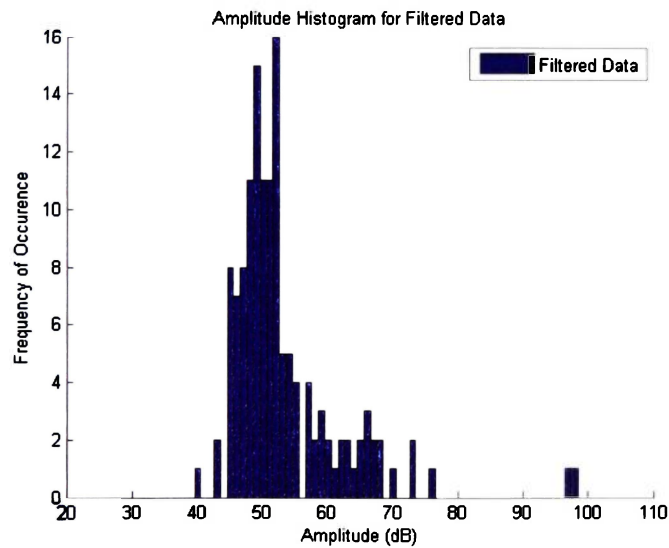
Figure 41: Final amplitude histogram for 7075-011

Figure 41 shows that Specimen 011 had multiple modes that were not discernible from

the noise until the data had been filtered. Two major modes at 50dB and 65dB became

58

apparent in the final amplitude histogram. Comparison of the 20x20 SOM data for 7075-011 (Figure 42) shows that although the network could run with the original data, the SOM was only able to shift the weights all the way to a value of -1, leaving very little structure in the network before the filtering technique had been applied. The filtered SOM was able to shift weights into two major and one minor grouping.
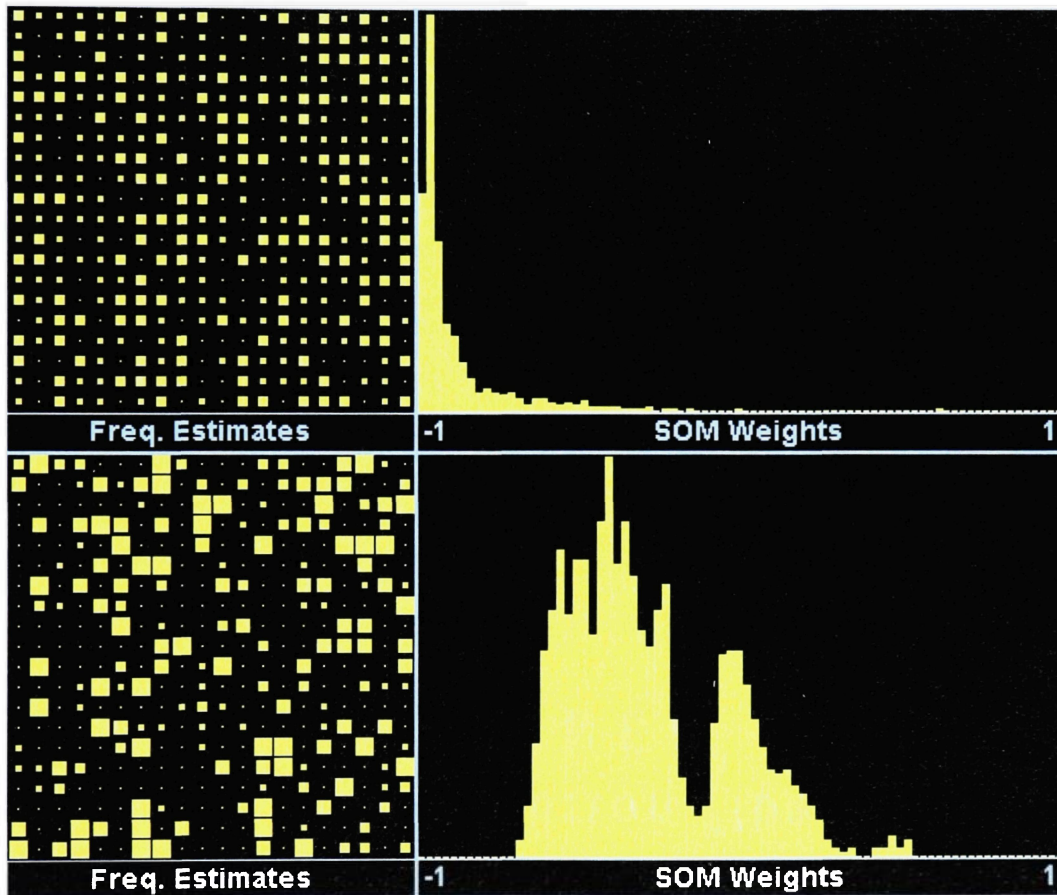
Figure 42: Unfiltered (top) and filtered (bottom) 7075-011 20x20 SOM

59

5.0 CONCLUSIONS

At this time, a number of conclusions can be drawn about the acoustic emission filtering technique and its effectiveness The first conclusion is that the technique appears to be a valid method for reduction of acoustic emission data The filter was able to reduce the data set by at least an order of magnitude in every case, and this significant reduction in data without a loss of defining characteristics can be a saver of both storage space and computing power None of the frequency components encountered after filtering exceeded 350kHz, so it is reasonable to conclude that EMI has successfully been removed Also, because a large majority of the data points removed by the filtering technique were diffuse over low frequency ranges, it can be concluded that the technique is also effective at removing mechanical noise

Secondly, as mentioned before, the defining characteristics of the failure mechanisms become more evident with a reduced data set The filter removes electromagnetic and mechanical noise, as well as multiple-hit data Elimination of these common types of noise leave a clearly defined signal extracted from the data, which then allows for easy visual inspection or mathematical and statistical analysis of the acoustic emission data

6.0 RECCOMMENDATIONS

Given that an effective filtering method exists that can be computerized, it is feasible to use it in a wide range of applications Real-time monitoring systems would be able to cycle through collected data and reduce it for more effective analysis Proof and ultimate strength testing involving acoustic emission could take much less time if the data were quickly reduced yet still carried a representative AE signal New handheld acoustic emission units gaining popularity in the field could be made more powerful with the addition of a fast, smart filter that requires minimal intervention by an untrained operator The key to obtaining such goals is further validation of this filtering technique on a wide variety of specimens Use on different materials and geometries would provide many different cases on which the use of the technique could be confirmed A database of ideal settings for the software for different setups should be compiled and made available to future users

7.0 REFERENCES

1. Hill, E. v.K., Israel, P L., and Knotts, G. L., "Neural Network Prediction of Aluminum-Lithium Weld Strengths From Acoustic Emission Amplitude Data," *Materials Evaluation*, September 1993, 1040-1051.

2. Hill, E. v.K., Walker, J. L., and Rowell, G. H., "Burst Pressure Prediction in Graphite/Epoxy Pressure Vessels Using Neural Networks and Acoustic Emission Amplitude Data," *Materials Evaluation*, June 1996, 744-754.

3. Faucett, L., Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1994.

4. Karl, J., Dion, S., and Spivey, N., "Neural Network Burst Pressure Predictions in Cryogenically Cooled Composite Overwrapped Pressure Vessels from Acoustic Emission Data," ASNT Spring 2006 Research Conference, American Society for Nondestructive Testing, Columbus, OH, 2006.

5. Ibekwe, E., "Neural Network Fatigue Life Prediction in 7075-T6 Aluminium from Acoustic Emission Data," MSAE Thesis, Embry-Riddle Aeronautical University, 2004.

6. Caudill, M. "Neural Networks Primer," *AI Expert*, November 1987, 46-52.

7. Miller, R. K., Hill, E. v.K., and Moore, P. (Eds.), Nondestructive Testing Handbook Volume 6: Acoustic Emission Testing, American Society for Nondestructive Testing, Columbus, OH, 2005.

8. Pollock, A. A., Practical Guide to Acoustic Emission Testing, Physical Acoustics Corporation, Princeton Junction, NJ, 1988.

9. Hill, E. v.K., "Predicting Burst Pressures in Filament-Wound Composite Pressure Vessels by Using Acoustic Emission Data," *Materials Evaluation*, December 1992, 1439-1445.

APPENDIX A: MATLAB CODE

```
%Import *.txt file into workspace as an 8 column array named "data"
%The array headers are as follows:
%point=data(:,1); press=data(:,2); ch=data(:,3); rise=data(:,4);
%counts=data(:,5); energy=data(:,6); dur=data(:,7); amp=data(:,8)


%composite specimen default
%  2% frequency extraction range,
%  700Hz frequencyrange
%  10 frequencies

%metal specimen default
%  4% frequency extraction range,
%  500Hz frequencyrange
%  10 frequencies



datacount=length(data);
data(:,9)=0;
data(:,9)=data(:,5)./(data(:,7)/1000000);   % conversion to Hz  (frequency)
data(:,10)=1:1:datacount;

%Set bounds for "obvious" frequency filter...
initlb=50000;  %approx top of auditory range
initub=2000000;  %2 mHz

frequencyrange=700; % set frequency range for distributions

freqstofind=10; % set number of frequencies to find

count=1;
newdatacount=1;
newdata=zeros(1,10);
highfreq=0;
%check to see if C/D is within the range, remove zero energy hits, read
%channel 1

%remember to store highest frequency
while count <= datacount
    if data(count,9) > initlb
        if data(count,9) < initub
            if data(count,6) > 0
```

```
              if data(count,7) < 100000
                if data(count,3) == 1
                newdata(newdatacount,:)=data(count,:);  %new data row
                newdatacount=newdatacount+1;   %increment counter
                newdata(newdatacount,:)=zeros;    %write zeros to next
                  if data(count,9)>highfreq
                      highfreq=data(count,9);   %store highest freq
                  end
                end
              end
            end
        end
      end
      count=count+1;
end


%show fundamental frequency distribution in bins with size determined by
%frequencyrange

bins=fix(highfreq/frequencyrange);
figure(1);
hold on
hist(newdata(:,9),bins);
axis tight
xlabel('Frequency (Hz)')
ylabel('Frequency of Occurence')
title('Frequency Distribution of Original Data')




%extract unsorted frequency histogram
unsortedfreqhist=hist(newdata(:,9),bins);
x=1;
while x <=length(unsortedfreqhist)
   unsortedfrequencies(x,1)=unsortedfreqhist(x);
   unsortedfrequencies(x,2)=x*frequencyrange;
x=x+1;
end

frequencies=flipud(sortrows(unsortedfrequencies,1));

maxhits = frequencies(1,2);
```

```
x=1;
while x <= freqstofind
fundfreq(x)=frequencies(x,2);
x=x+1;
end


%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%  OK, now we analyze 2% in either direction of the
maxfreqs
%%%%%%%%%%%%%%%%%%%%%%%

%isolate fundamental frequencies
fundcount=1;
finaldatacount=1;
finaldata=zeros(1,10);

while fundcount <=length(fundfreq)

lb=fundfreq(fundcount)-(highfreq/50);
ub=fundfreq(fundcount)+(highfreq/50);

count=1;
while count <= length(newdata)
    if newdata(count,9) > lb
        if newdata(count,9) < ub
            finaldata(finaldatacount,:) = newdata(count,:);
            finaldatacount=finaldatacount+1;
        end
    end
count=count+1;
end

fundcount=fundcount+1;
end

count=1;
cleancount=1;
while count <= length(finaldata)
    if finaldata(count, 1) > 0
        cleandata(cleancount,:)=finaldata(count, :);
        cleancount=cleancount+1;
    end
count=count+1;
```

end

%------------make D v C plot---------------
figure(2);
plot(data(:,5),data(:,7),' r', 'MarkerSize', 1);
hold all
plot(cleandata(:,5),cleandata(:,7), 'o b', 'MarkerSize', 2);
legend('Raw Data','Filtered Data')
xlabel('Counts')
ylabel('Duration(microseconds)')
title('Duration vs. Counts for Filtered and Raw Data')

%----------make comparative amplitude histograms------------
figure(3);
hold on
hist(data(:,8),60:100)
h = findobj(gca,'Type','patch');
set(h,'FaceColor','r')
hist(cleandata(:,8),60:100)
legend('Raw Data','Filtered Data')
xlabel('Amplitude (dB)')
ylabel('Frequency of Occurence')
title('Amplitude Histogram for Filtered and Raw Data')

%----------make final amplitude histogram------------
figure(4);
hold on
set(h,'FaceColor','r')
hist(cleandata(:,8),60:100)
legend('Filtered Data')
xlabel('Amplitude (dB)')
ylabel('Frequency of Occurence')
title('Amplitude Histogram for Filtered Data')

%%%
%%%%%%%%%%%%%%%%%%%%%%
%%υ ο%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%

%%% calculate curve fitting moments

histogram=hist(cleandata(:,8),41);

distribution=transpose(histogram);

%break;

%compute mu, the mean%
mu=mean(distribution)

%compute sigma, the standard deviation%
sigma=std(distribution)

%compute m2, the second moment of the distribution
m2=sigma^2

% this will give you the number of rows:
n = size(distribution,1);

m3loop=0; m3temp=0;
%compute m3, the third moment of the distribution
for i=1:n
  m3temp=(distribution(i)-mu)^3;
  m3loop=m3loop+m3temp;
end
m3=m3loop*(1/n)

m4loop=0; m4temp=0;
%compute m4, the fourth moment of the distribution
for i=1:n
  m4temp=(distribution(i)-mu)^4;
  m4loop=m4loop+m4temp;
end
m4=m4loop*(1/n)

%compute beta1 (relative skewness squared)

67

```
beta1=(m3/((m2)^(3/2)))^2

%compute beta2 (relative kurtosis)
beta2=m4/((m2)^2)
```

COPV Specimen 002

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

70

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 009	

71

COPV Specimen 010

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 013

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 014

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 018

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

75

COPV Specimen 020

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 025

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

COPV Specimen 026

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

78

7075 Specimen 001

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 002

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 003

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

Frequency Distribution of Original Data

Amplitude Histogram for Filtered and Raw Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 005

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 006

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 007

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 008

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 009

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

87

7075 Specimen 010

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

7075 Specimen 011

Frequency Distribution of Original Data

Duration vs Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data

90

Frequency Distribution of Original Data

Frequency of Occurence

Frequency (Hz)

Duration vs. Counts for Filtered and Raw Data

Duration(microseconds)

Counts

Raw Data
Filtered Data

Amplitude Histogram for Filtered and Raw Data

Frequency of Occurence

Amplitude (dB)

Raw Data
Filtered Data

Amplitude Histogram for Filtered Data

Frequency of Occurence

Amplitude (dB)

Filtered Data

7075 Specimen 013

Frequency Distribution of Original Data

Duration vs. Counts for Filtered and Raw Data

Amplitude Histogram for Filtered and Raw Data

Amplitude Histogram for Filtered Data
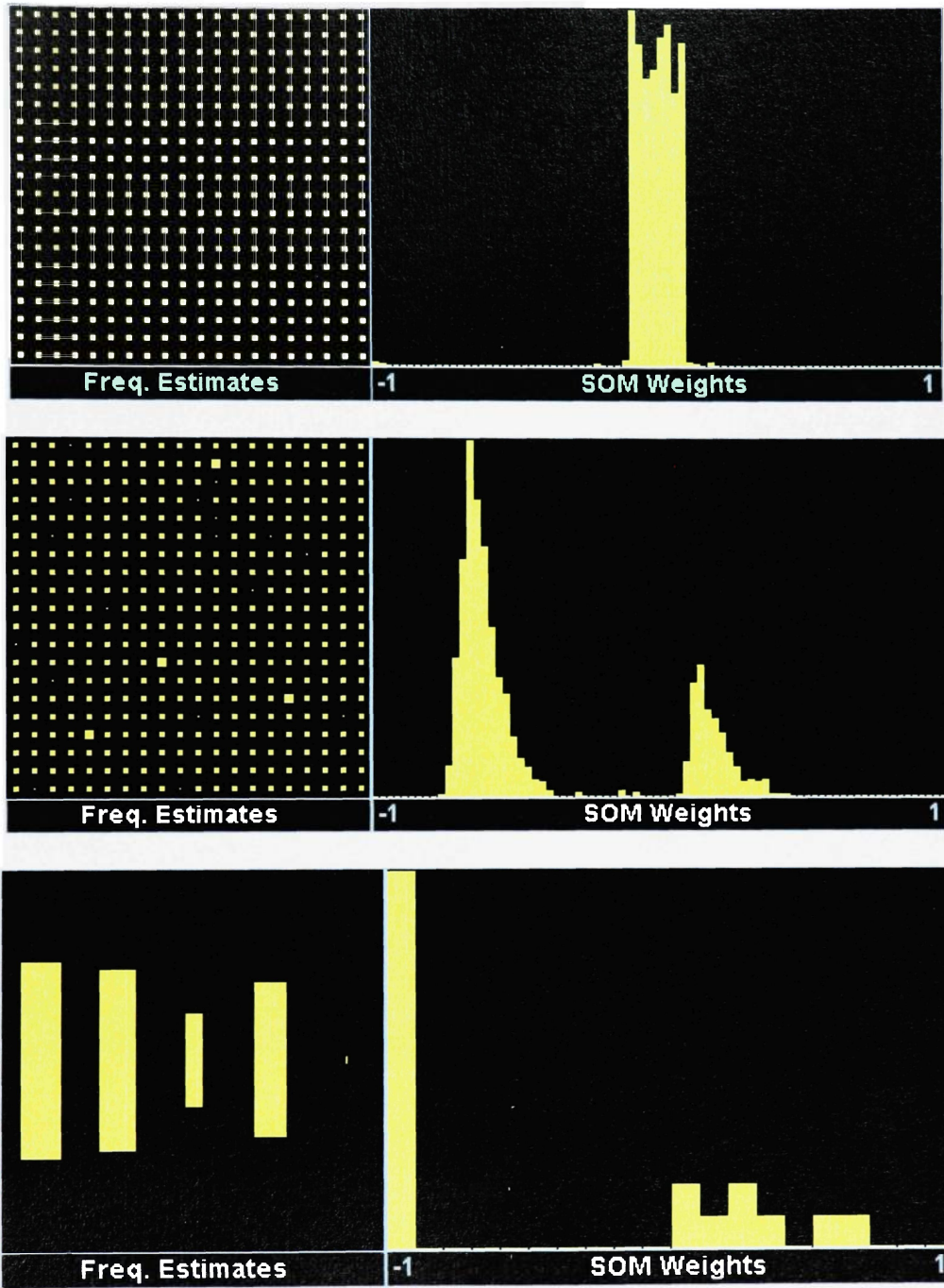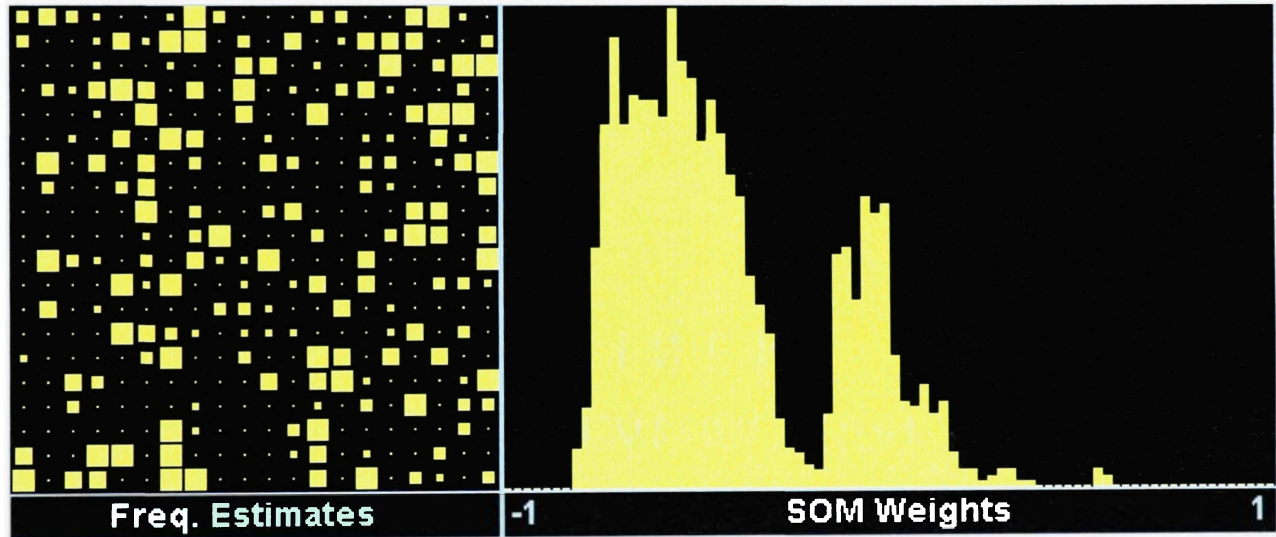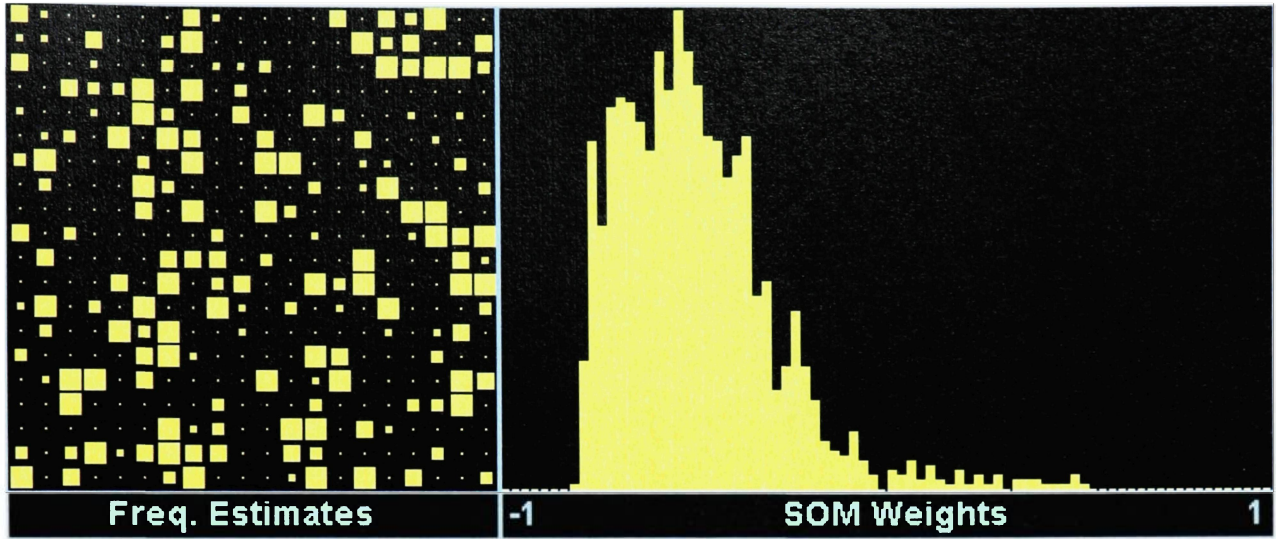
APPENDIX C: SELF-ORGANIZING MAP DATA

COPV Specimen 002

COPV Specimen 003

COPV Specimen 009

COPV Specimen 010

Freq. Estimates | SOM Weights

95

COPV Specimen 013

Freq. Estimates | SOM Weights (-1 to 1)

COPV Specimen 014
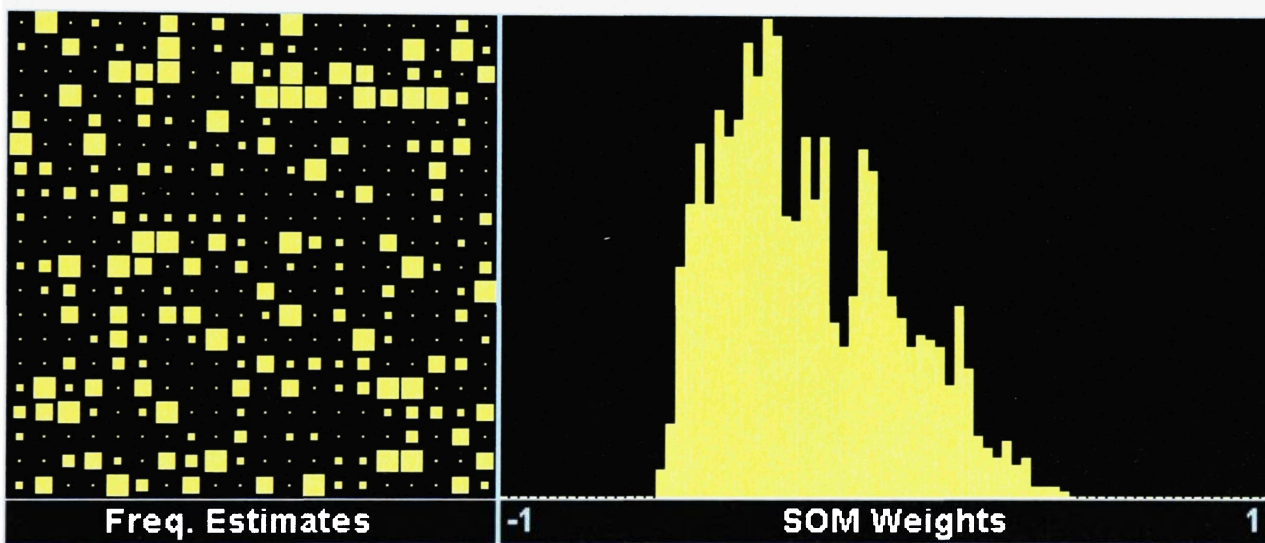
COPV Specimen 018

98

COPV Specimen 020

COPV Specimen 025

COPV Specimen 026

101

7075 Specimen 001
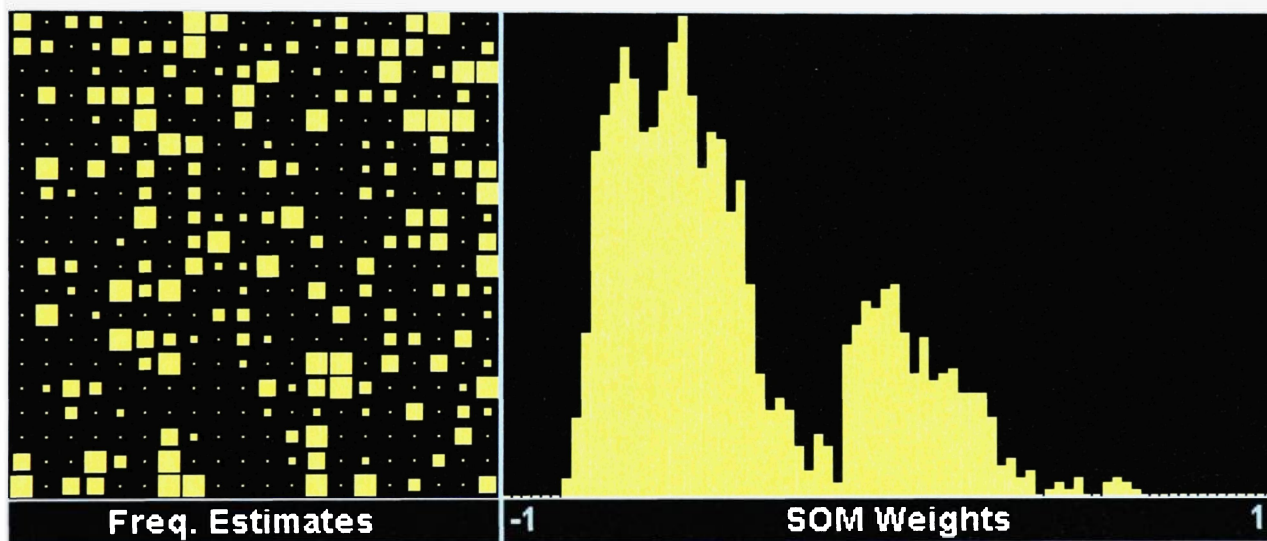

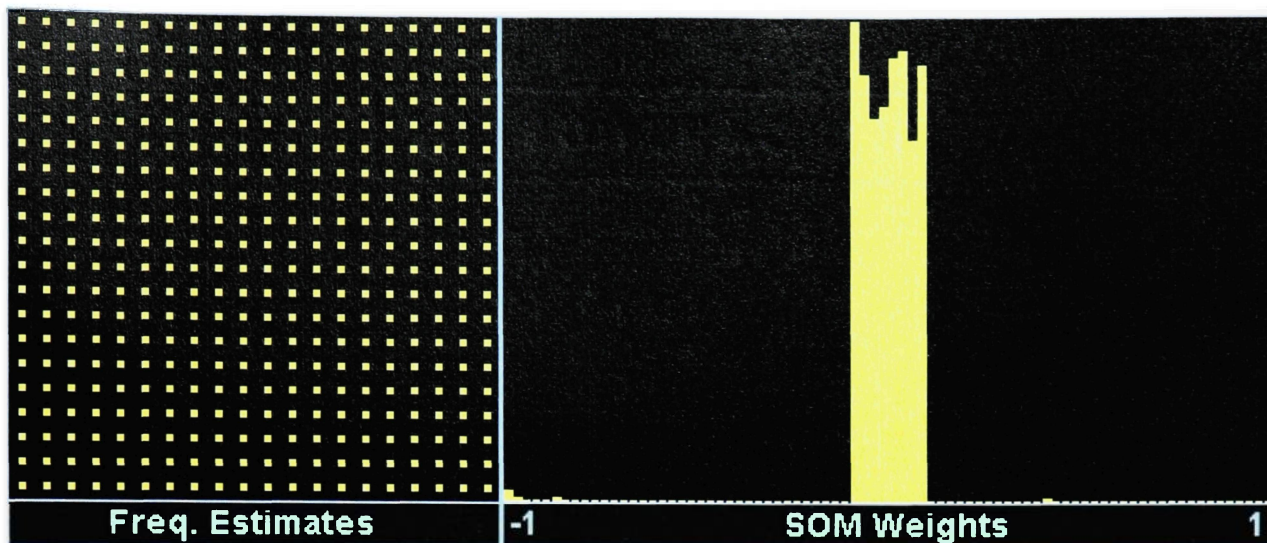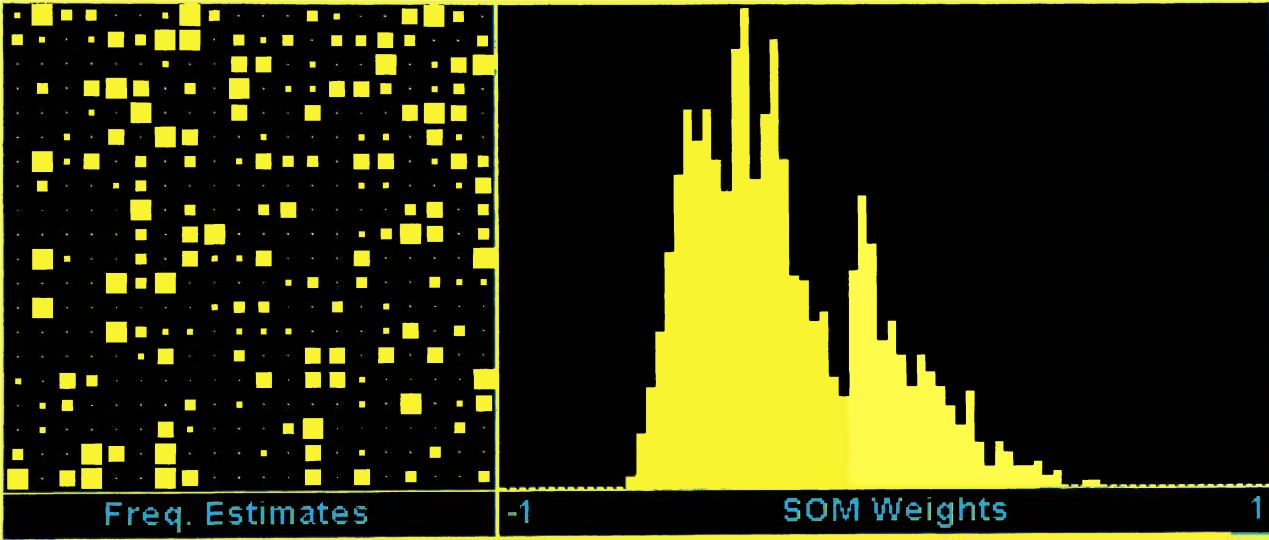102
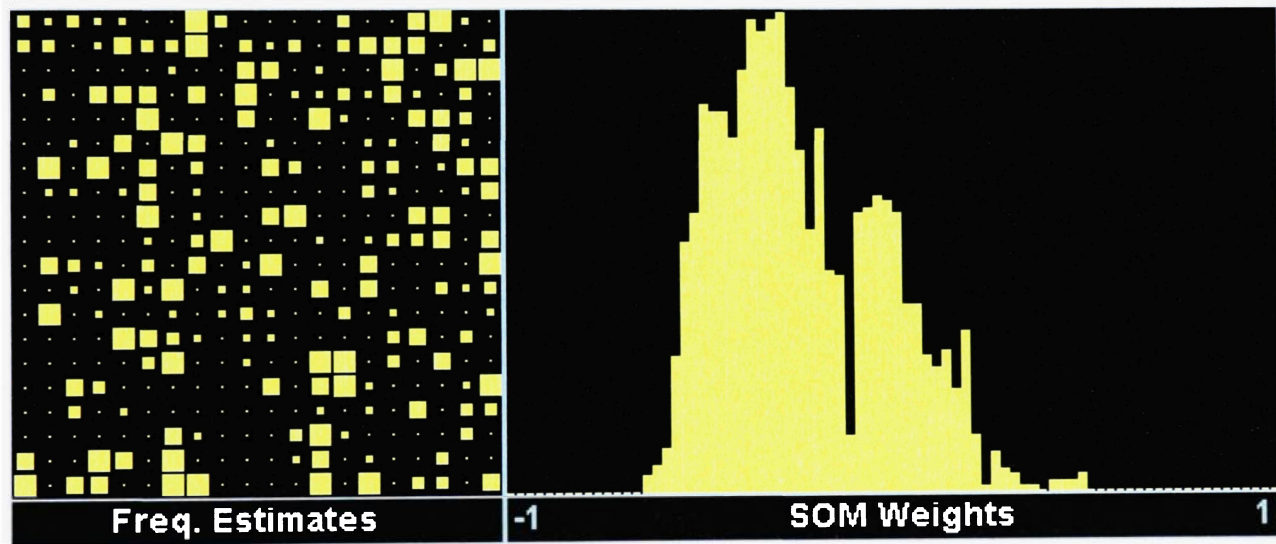
7075 Specimen 002

7075 Specimen 003
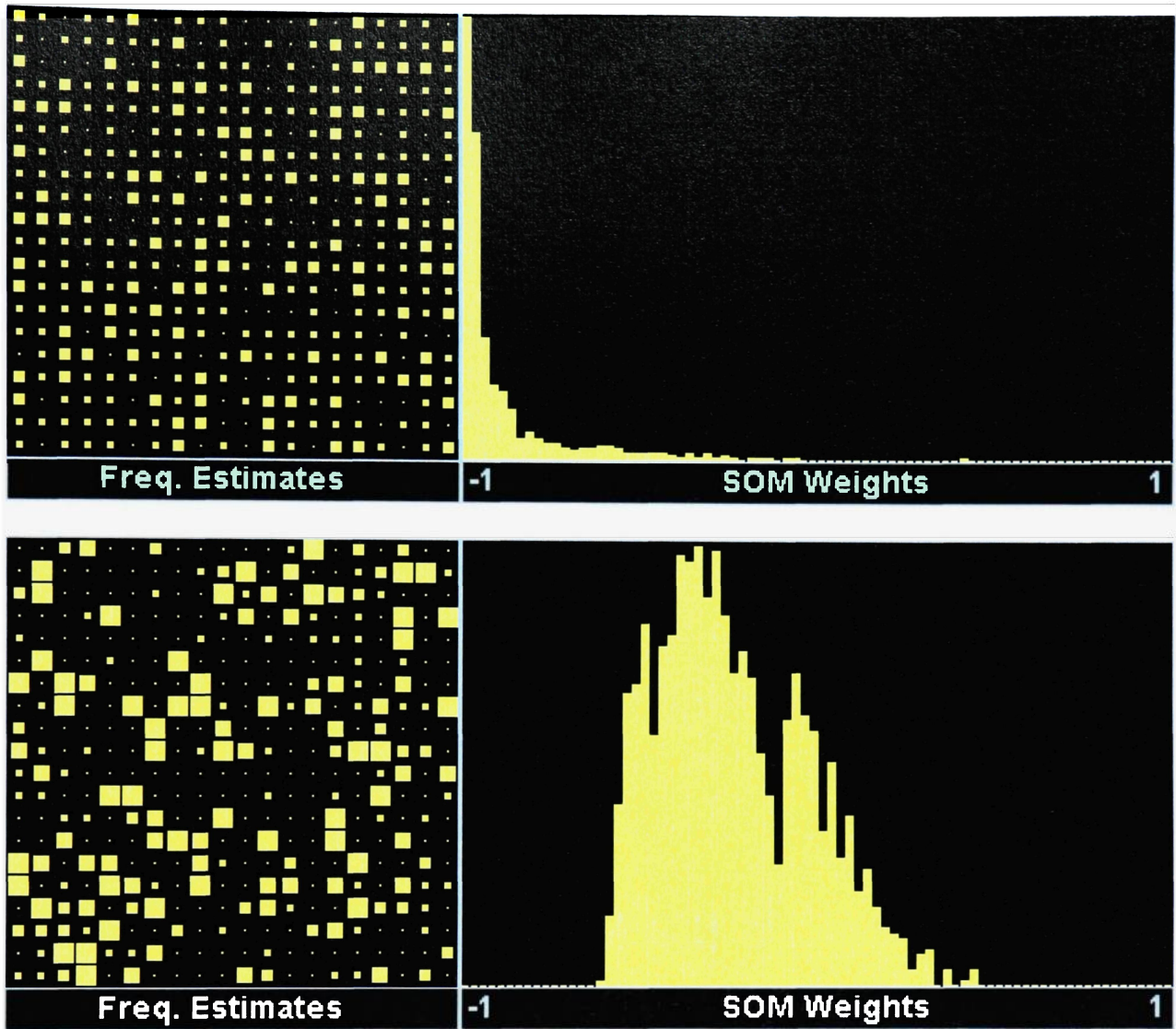

104

7075 Specimen 004

105

7075 Specimen 005

7075 Specimen 007

108

7075 Specimen 008

109

7075 Specimen 009

110
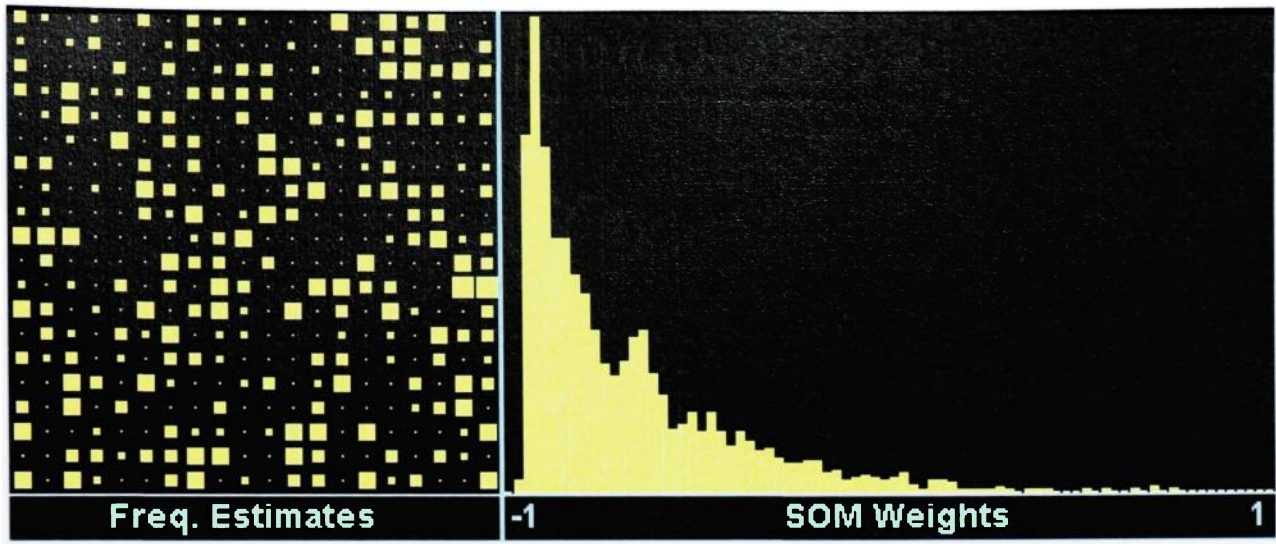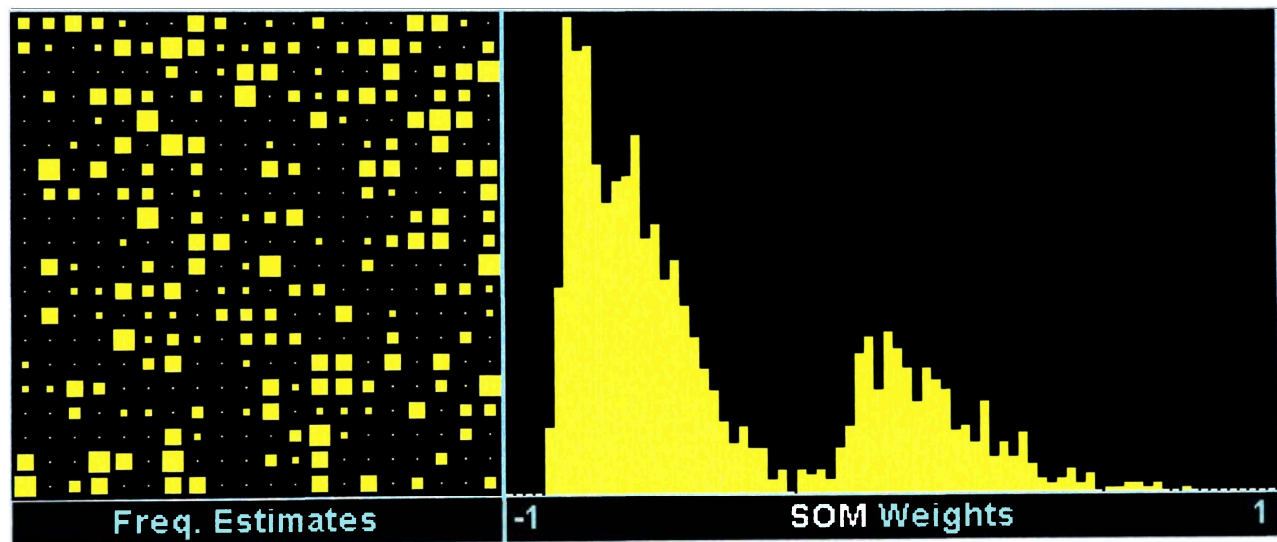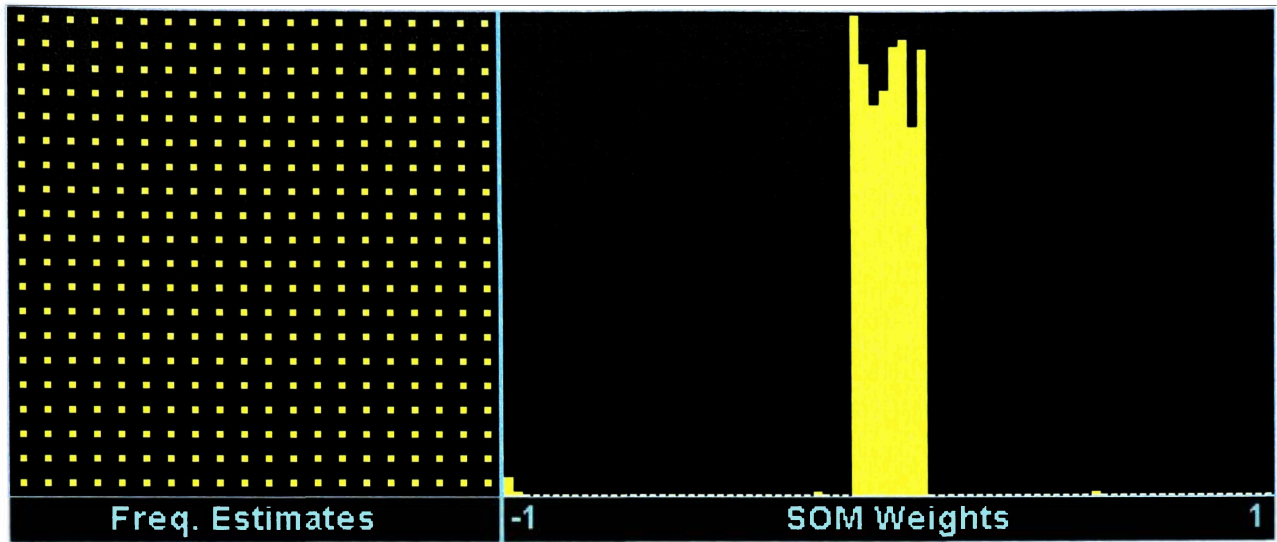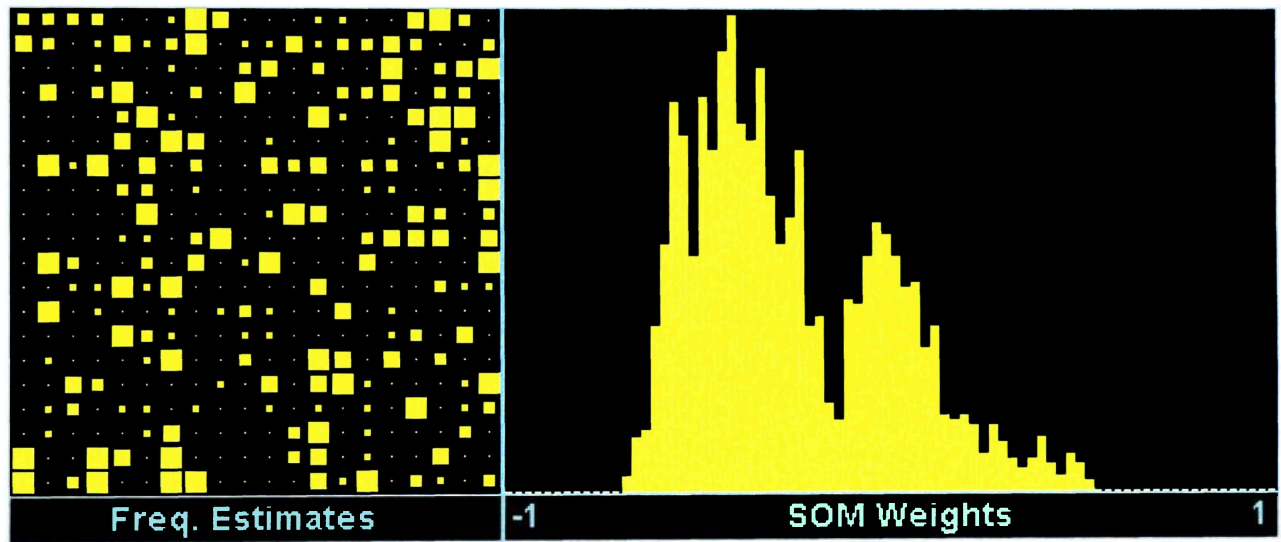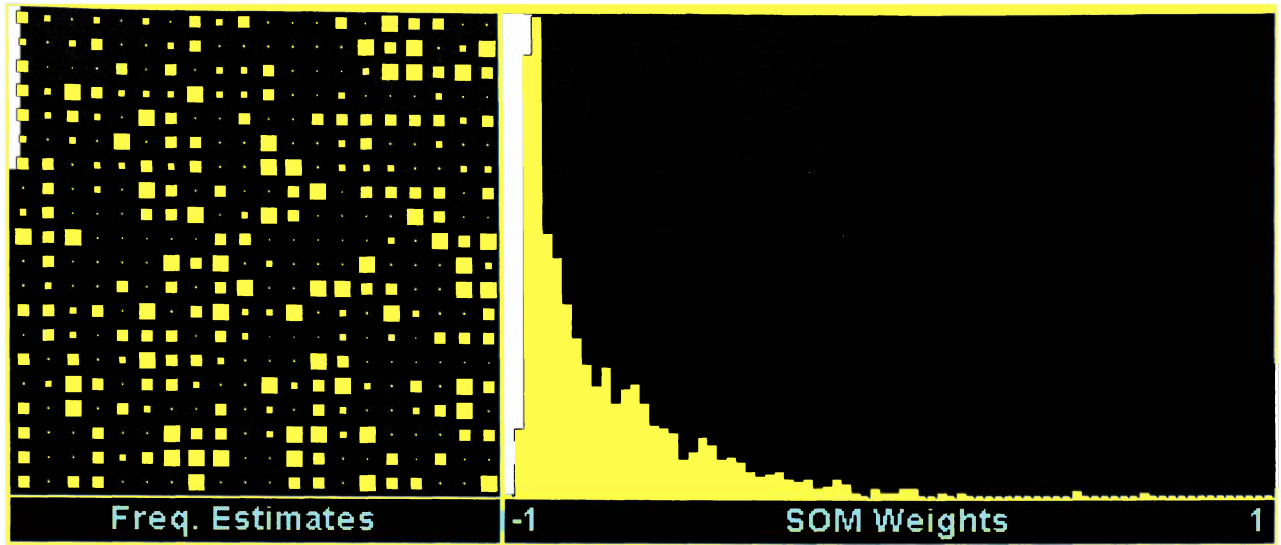
7075 Specimen 010

7075 Specimen 011
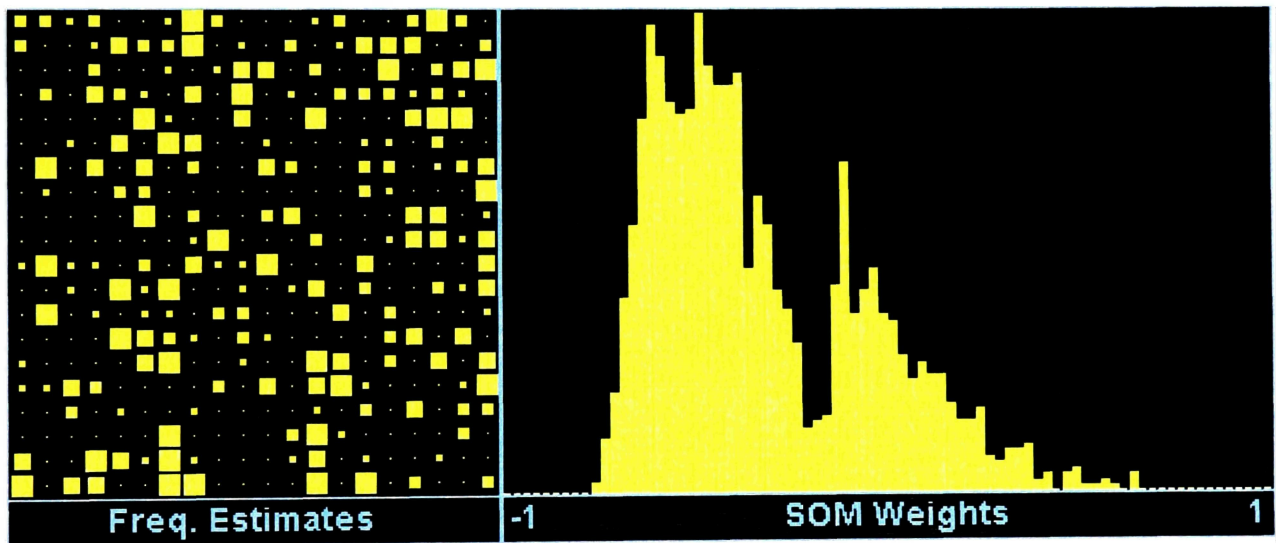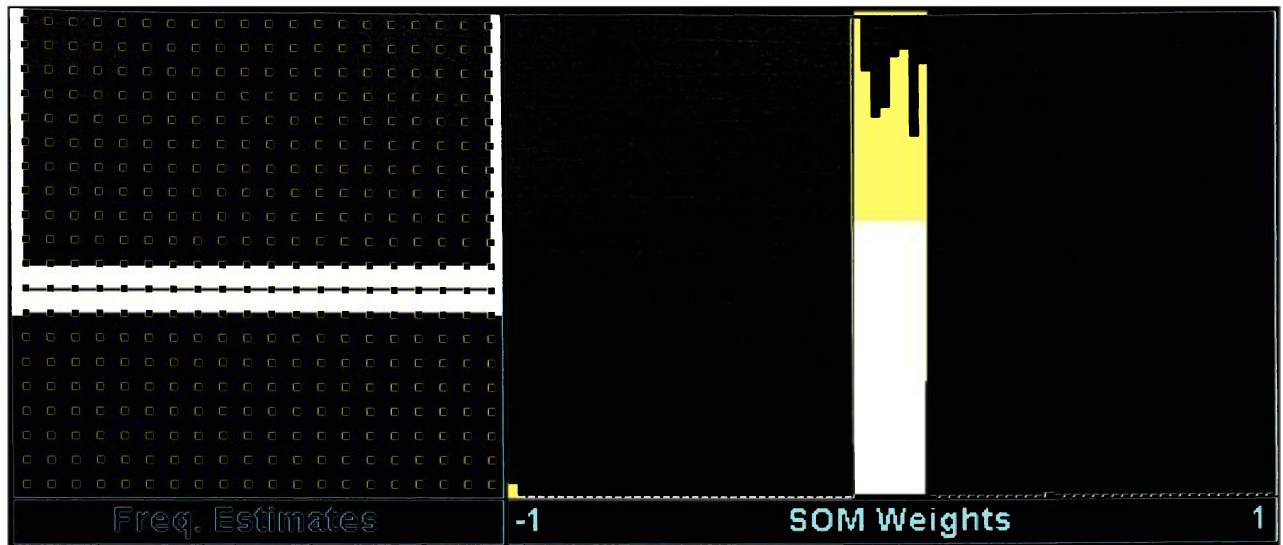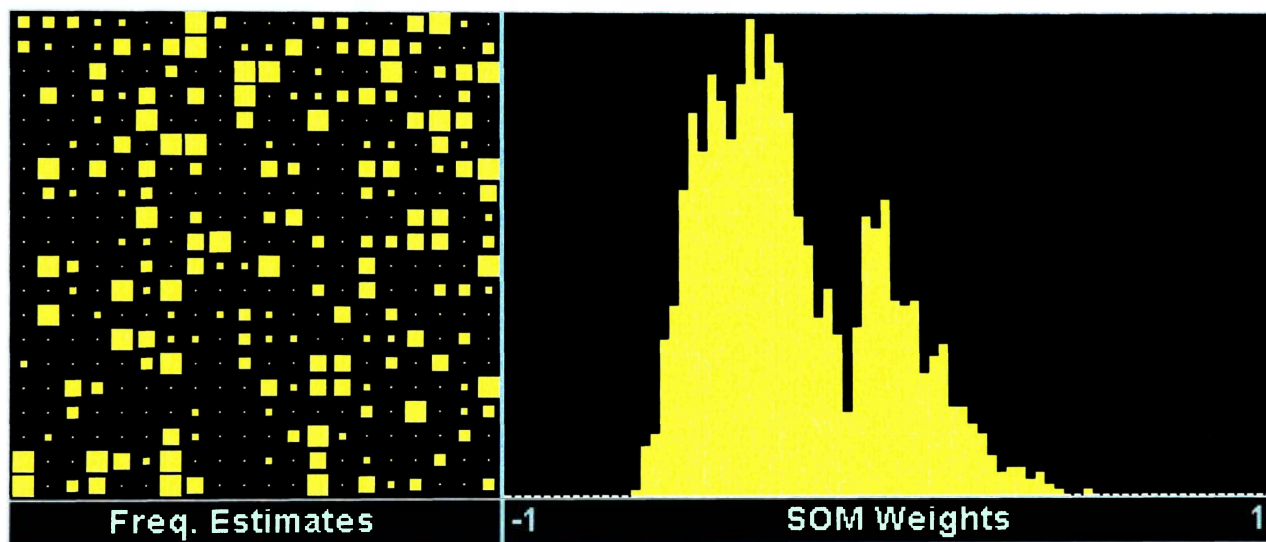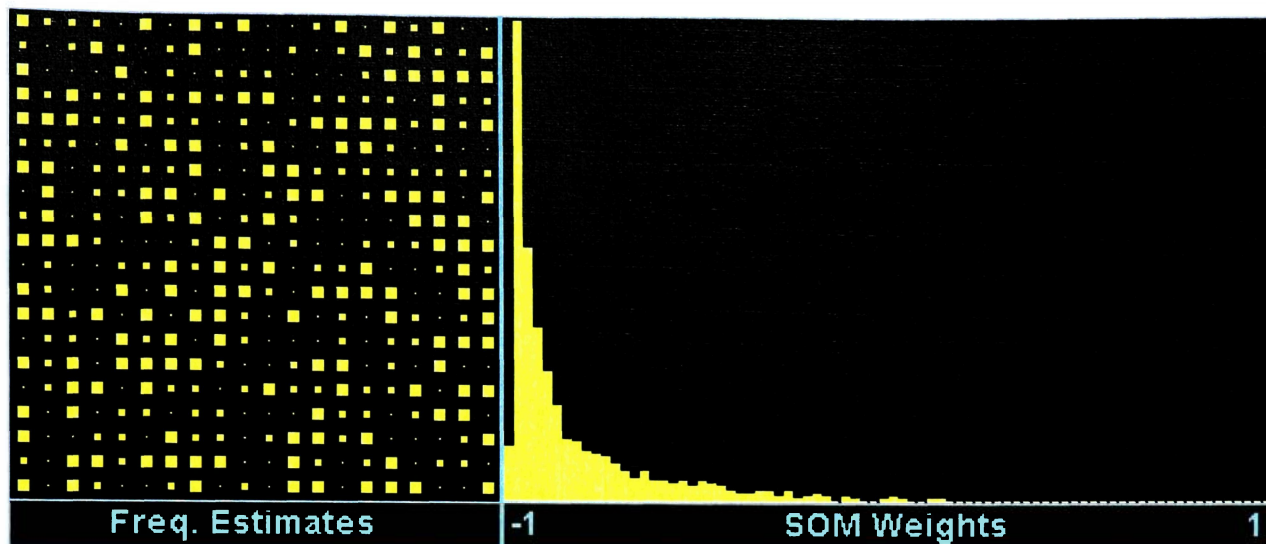

112

7075 Specimen 012

113

7075 Specimen 013

114
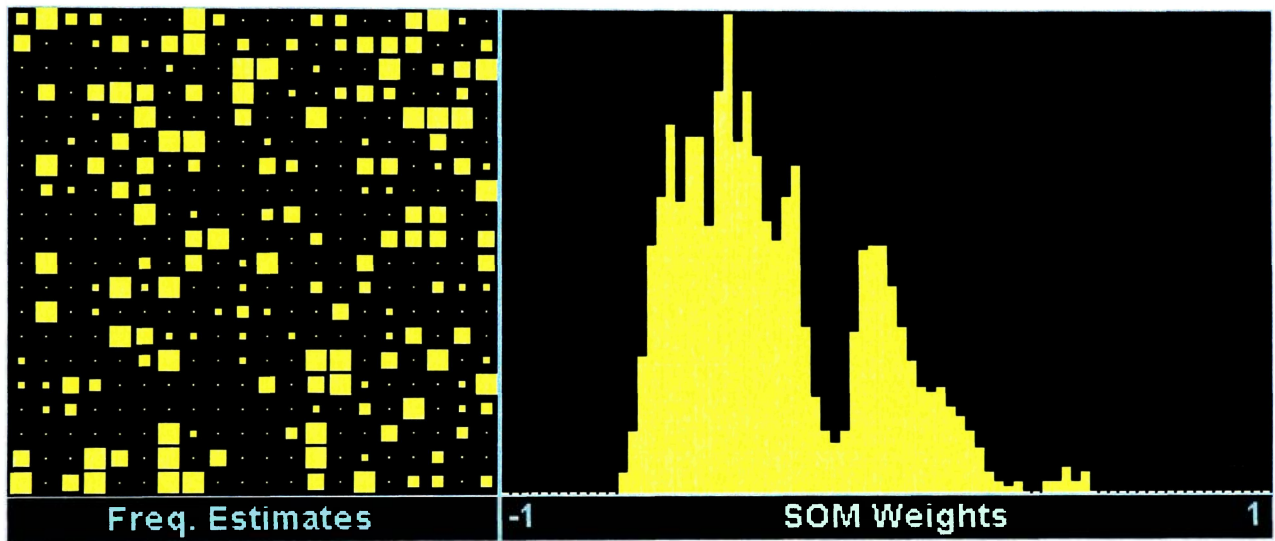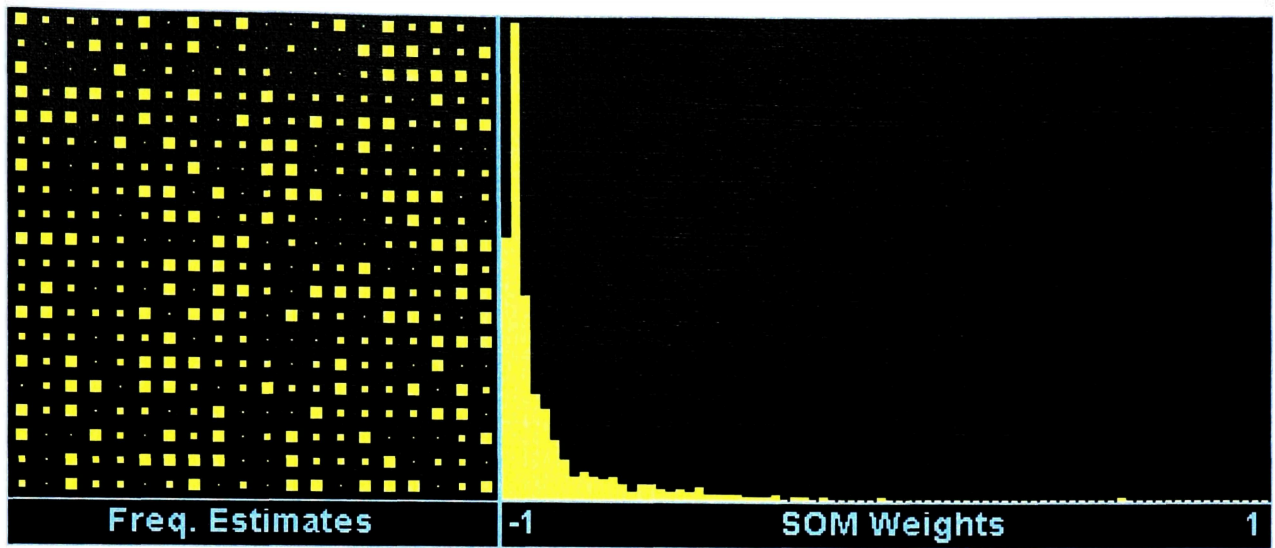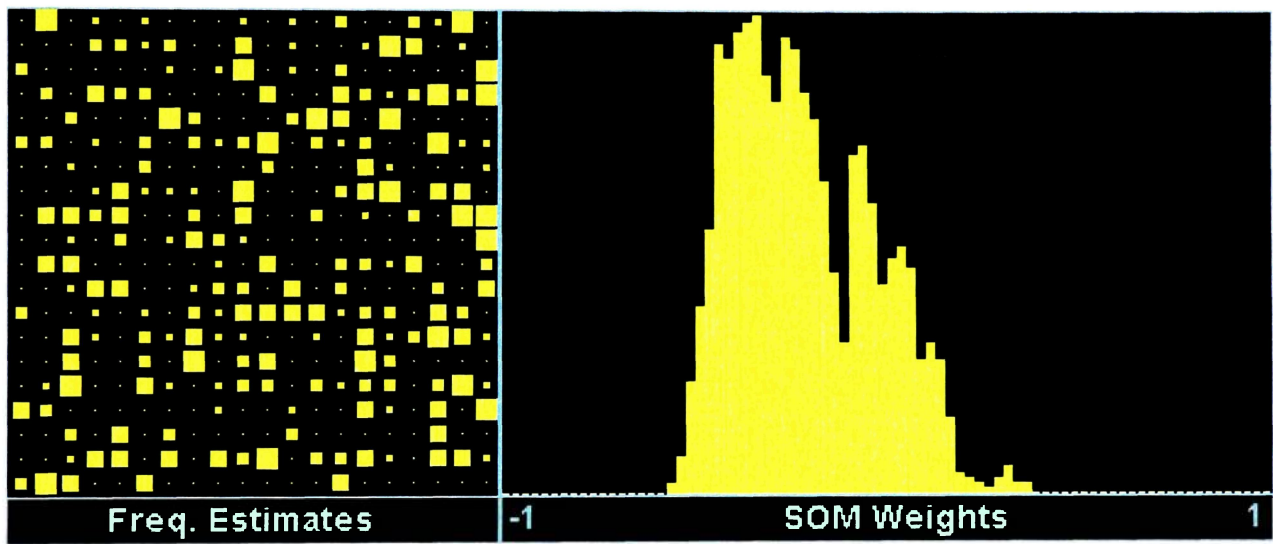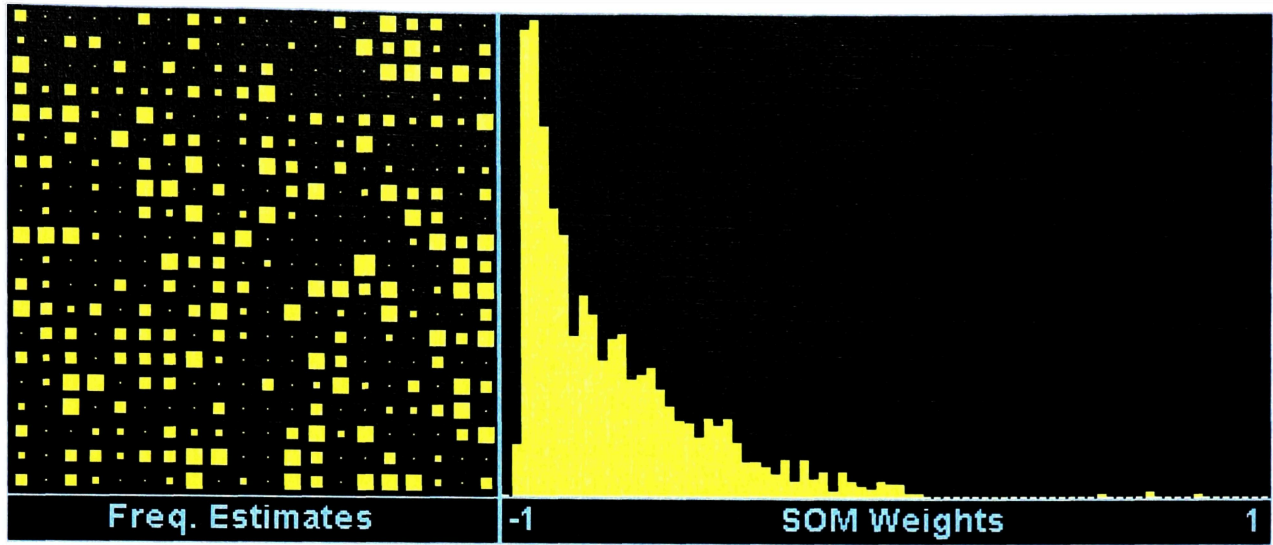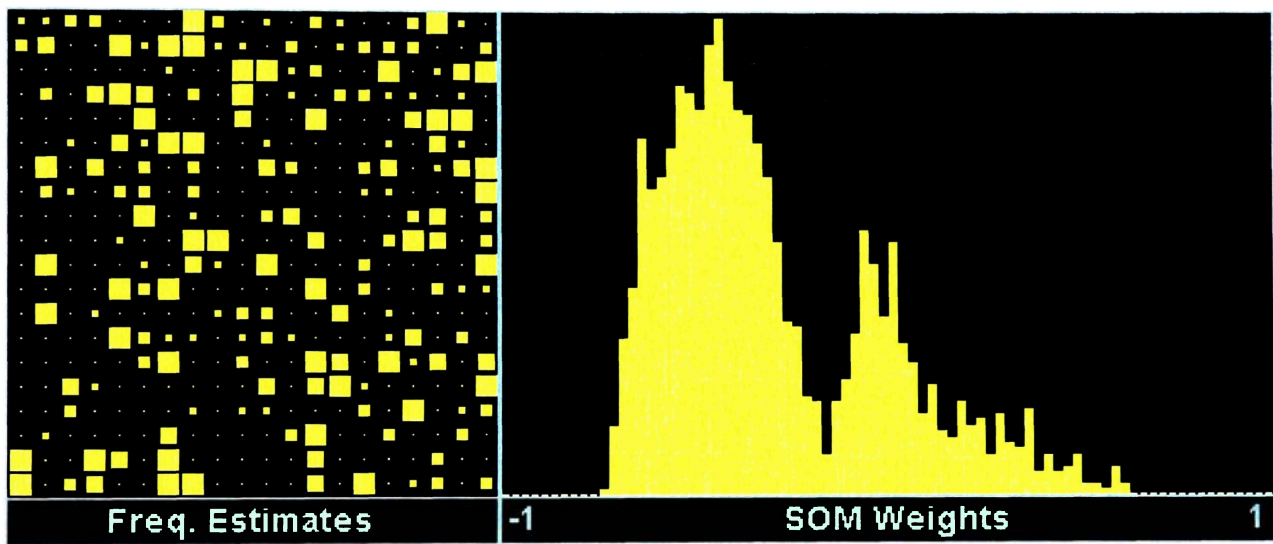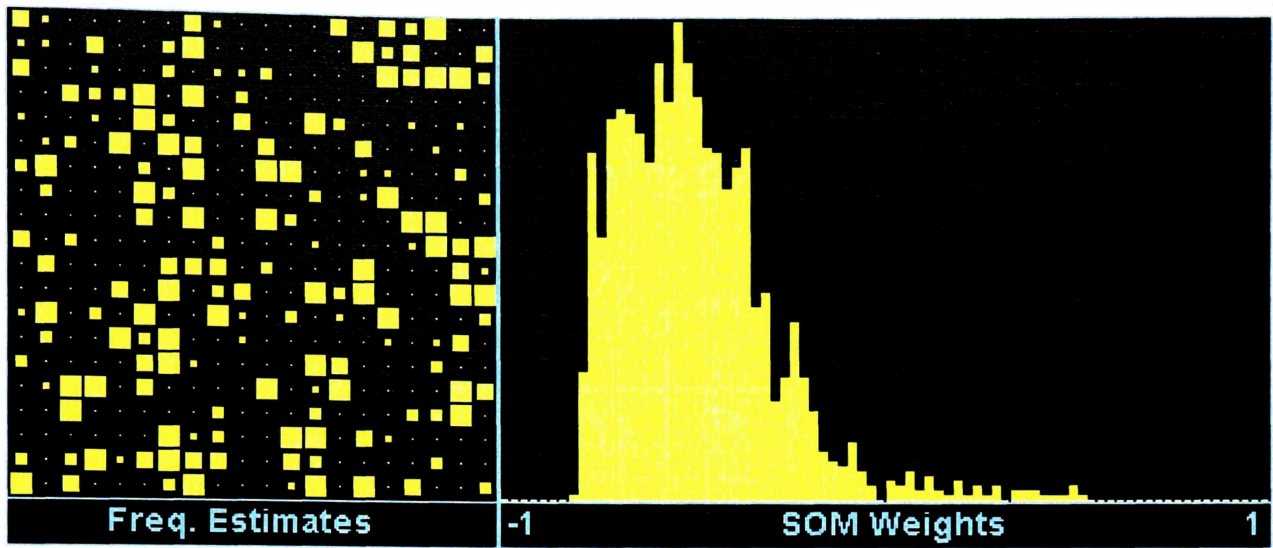
APPENDIX D: BACKPROPAGATION NEURAL NETWORK THEORY

The mathematical theory behind the application of backpropagation neural networks (BPNNs) is complicated, but to gain a basic understanding, a piece-by-piece look at the BPNNs used in this research will be explored. BPNNs are an n-dimensional vector-based mathematical network that is comprised of three groupings of nodes (neurons). The first grouping is referred to as the "input layer", the second as the "hidden layer" (of which there may be more than one), and the third is the "output layer".

The input layer in this research consisted of a 44x1 dimensional vector comprised of the following entries: (1) a three-variable categorical variable set, (arranged as 1,0,0 for undamaged COPVs, 0,1,0 for impact damaged COPVs, and 0,0,1 for lacerated COPV specimens) and (2) the 41 integer variables representing the frequency of event occurrence at amplitudes from 60dB to 100dB that occurred during each acoustic emission test.

The number of neurons in the hidden, or mapping layer (this network used a single hidden layer -- a common occurrence) was 11. The number of nodes in the hidden layer correlates directly to the closeness of fit for the network training data. A low number of neurons would not fit the testing data well, while a high number of neurons would fit the data too closely, resulting in the network not predicting well when it encounters the test data. A goal in constructing and training a prediction-type BPNN is to change the

number of neurons in the hidden layer until the test data and training data error levels are within the same order of magnitude At this point the network has been properly trained Another approach involves matching the number of expected mechanisms in the material with a set of two neurons each It has been shown from a geometric perspective that two neurons are required to model the humps in the amplitude histogram data One can visualize the way this works by drawing a complex curve representing the histogram values and then attempting to fit it with a series of straight lines oriented at different locations and angles In this example, each straight line correlates to a hidden layer neuron

The final layer is the output layer, which consisted of a single value that is dependent on the total of all of the mathematical interactions in the network – in this research, the output value was burst pressure (in psi) While the network is training, the weight of the connections between the input layer nodes and the hidden layer nodes as well as the hidden layer nodes and the output node are continuously updated and shifted to produce a result that is further from the initial value and closer to the known (as tested) output value When the network meets the convergence criterion set by the user, it is said that it has successfully been trained to recognize the pattern of inputs and the appropriate output

At this point the network user runs one more network iteration -- this time with a set of inputs that is unfamiliar to the network Based on its newly acquired ability to recognize

116

patterns in the similarly structured amplitude histogram data from training, the network

produces a predicted burst pressure as the output.

APPENDIX E: SELF-ORGANIZING MAP THEORY

This brief tutorial aims to teach the reader the self-organizing map and the application of SOMs to a simple (hand calculated) problem

Teuvo Kohonen developed the SOM in the early 80s One of his first applications of a SOM was to convert spoken words into computerized text The SOM is a dual layer (input layer and PE layer), unsupervised, competitive network The network learns by selectively allowing PEs to learn by minimizing the linear geometric distance (Euclidean distance) between the input and the weight vectors. The determination of which PEs learn is controlled by a user-set neighborhood factor. Additionally, the neighborhood factor allows any PEs within that neighborhood of the winning PE to learn

SOM networks are trained using an iterative process Starting with an initially-random set of weights, the algorithm gradually adjusts them to reflect the clustering of the training data The Kohonen algorithm iterates a given number of epochs, on each epoch executing each training case and selecting the winning PE, then adjusting its weights and the weights of the PEs in the winner s neighborhood

Since the SOM model does not have an output it is a unique system. At a cursory glance it might appear to be of little value, however, that assessment is incorrect. SOM networks learn the structure of the data, and are therefore excellent at pattern recognition and exploratory data analysis

It is possible to view the results by outputting the n-dimensional weight data onto a two dimensional space. However, the detail in the data is lost and cannot be recovered in 2D. For example, a 3D image can be viewed on a 2D surface, but only the information about the visible plane is available. Therefore, using the same logic, a SOM can display n-dimensional data on a 2D plane, where clustering and classification will be evident, though complex inferences about the data cannot be made unless it is compared to the original data set.
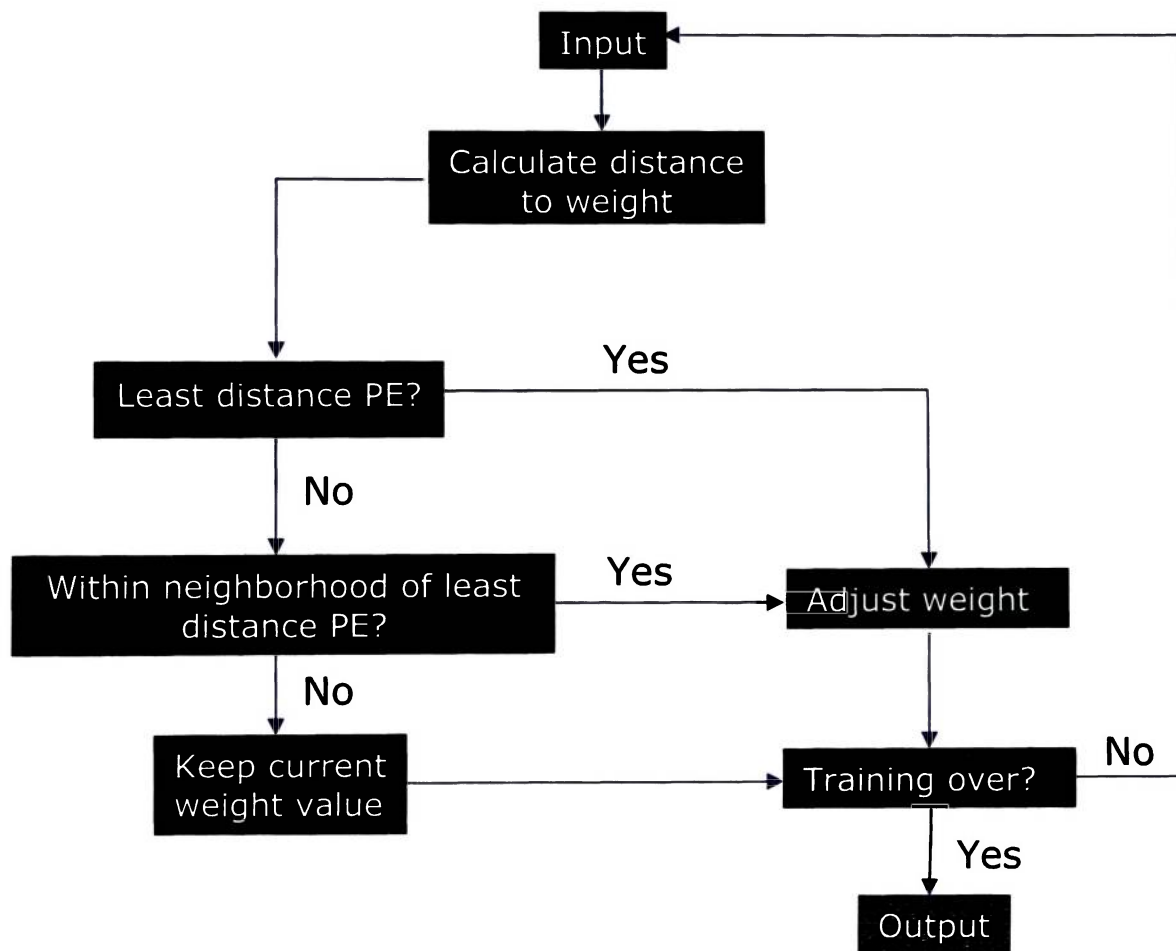
Figure 43: SOM algorithm flow
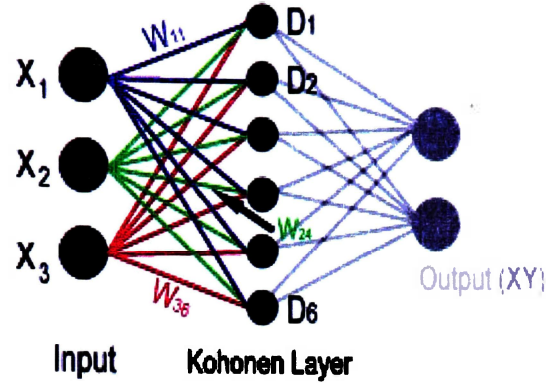
Figure 44: Kohonen architecture with forced X-Y output

In order to better understand how the network actually works, it is important to do an example that can be worked out by hand. The example in this section can be done with simple hand calculations in a few minutes. Consider the following Kohonen self-organizing network:

- 6 Neurons

- Input vector: $X_{input} = \begin{bmatrix} 0.700 & 0.900 & 0.200 \end{bmatrix}$

- Learning coefficient: $\alpha = 0.15$

- Initial weights:

$$W_{ij} = \begin{bmatrix} 0.300 & 0.600 & 0.100 & 0.400 & 0.800 & 0.100 \\ 0.700 & 0.100 & 0.400 & 0.300 & 0.900 & 0.800 \\ 0.100 & 0.200 & 0.700 & 0.600 & 0.100 & 0.200 \end{bmatrix}$$

These values of W_{ij} were randomly generated and rounded to the nearest tenth for ease of hand manipulation. The first step in the process is to determine the Euclidean distances between the input PEs and the individual neurons in the computing layer.

120

$$D = \sqrt{(W_{ij} - X_{input})^2}$$

$$D(1) = \sqrt{(w_{11} - x_1)^2 + (w_{21} - x_2)^2 + (w_{31} - x_3)^2} = 0.458$$

$$D(2) = \sqrt{(w_{12} - x_1)^2 + (w_{22} - x_2)^2 + (w_{32} - x_3)^2} = 0.806$$

$$D(3) = \sqrt{(w_{13} - x_1)^2 + (w_{23} - x_2)^2 + (w_{33} - x_3)^2} = 0.927$$

$$D(4) = \sqrt{(w_{14} - x_1)^2 + (w_{24} - x_2)^2 + (w_{34} - x_3)^2} = 0.781$$

$$D(5) = \sqrt{(w_{15} - x_1)^2 + (w_{25} - x_2)^2 + (w_{35} - x_3)^2} = 0.141$$

$$D(6) = \sqrt{(w_{16} - x_1)^2 + (w_{26} - x_2)^2 + (w_{36} - x_3)^2} = 0.608$$

Equation 1: Calculating Euclidean distances

In this case, $D(5)$ has the distance closest to zero. With a neighborhood factor of 1, that means that the winning neuron and the neighbors within a range of 1 neuron, $D(4)$ and $D(6)$, will undergo the learning process.

The neurons that undergo learning will have their weights updated with the following equation:

$$W_{ij(new)} = W_{ij(old)} + \alpha(X_i - W_{ij(old)})$$

Equation 2: SOM weight learning

Here α is the learning coefficient, a quantity that determines how fast the learning neurons will approach the quantities we are sorting for. A high α value can cause overshoot, like an underdamped oscillator, while a low α value may cause very slow learning, like an overdamped oscillator.

121

The calculations of the new network weights yield:

$$W_{u-NEW} = \begin{bmatrix} 0.300 & 0.600 & 0.100 & 0.550 & 0.750 & 0.400 \\ 0.700 & 0.100 & 0.400 & 0.600 & 0.900 & 0.850 \\ 0.100 & 0.200 & 0.700 & 0.400 & 0.150 & 0.200 \end{bmatrix}$$

The values in light red are the weights connected to the winning neuron. The dark-red values are the neighbors.

Notice that after running the data through the network, the weights of the winning and neighboring PEs have "trained" or "self-organized" to be closer to the input vector's values. After five iterations, the weights have shifted to the following values:

$$W_{u-ITR5} = \begin{bmatrix} 0.300 & 0.600 & 0.100 & 0.650 & 0.720 & 0.600 \\ 0.700 & 0.100 & 0.400 & 0.800 & 0.900 & 0.880 \\ 0.100 & 0.200 & 0.700 & 0.260 & 0.180 & 0.200 \end{bmatrix}$$

After 10 Iterations:

$$W_{u-ITR10} = \begin{bmatrix} 0.300 & 0.600 & 0.100 & 0.680 & 0.710 & 0.660 \\ 0.700 & 0.100 & 0.400 & 0.860 & 0.900 & 0.890 \\ 0.100 & 0.200 & 0.700 & 0.230 & 0.190 & 0.200 \end{bmatrix}$$

After 20 Iterations:

$$W_{u-ITR20} = \begin{bmatrix} 0.300 & 0.600 & 0.100 & 0.700 & 0.700 & 0.700 \\ 0.700 & 0.100 & 0.400 & 0.900 & 0.900 & 0.900 \\ 0.100 & 0.200 & 0.700 & 0.200 & 0.200 & 0.200 \end{bmatrix}$$

It becomes evident that W_{14}, W_{15}, and W_{16} have shifted toward 0.700. W_{24}, W_{25}, and W_{26} have shifted toward 0.900, and W_{34}, W_{35}, and W_{36} have shifted toward 0.200. If multiple input vectors had been used to train the network, weights would have shifted toward clusters of the input values. Essentially, the SOM would be finding correlations between the input vectors and organizing them together in the network weights.

APPENDIX F: NEURALWORKS SOFTWARE SETUP
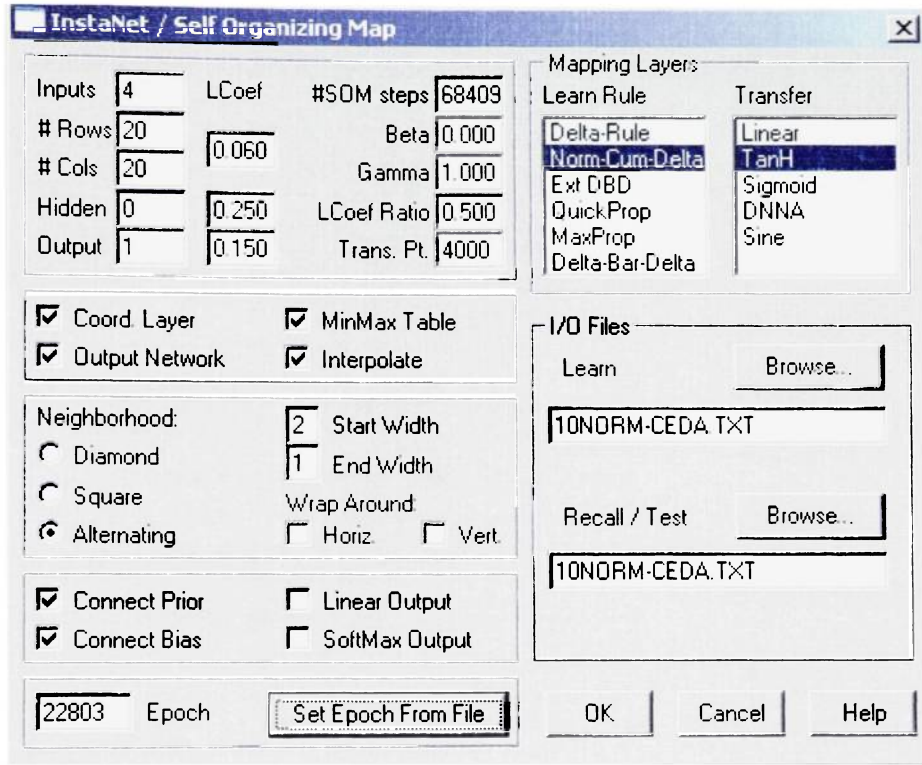
SELF-ORGANIZING MAP SETUP:

Figure 45: NeuralWorks setup screen for SOM network

Figure 45 shows the setup screen for the SOM networks utilized in this research. Four inputs (counts, energy, duration, and amplitude parameters) were used in a 20x20 network in this case. There is no hidden layer, and the output is arbitrary. SOM steps and epoch are determined automatically by the software based on specimen data file size. Neighborhood size is 2 before transition and 1 after transition. Transition point was set to

4000 because many of the data files were very small. A normalized cumulative delta learning rule was used with a hyperbolic tangent transfer function.
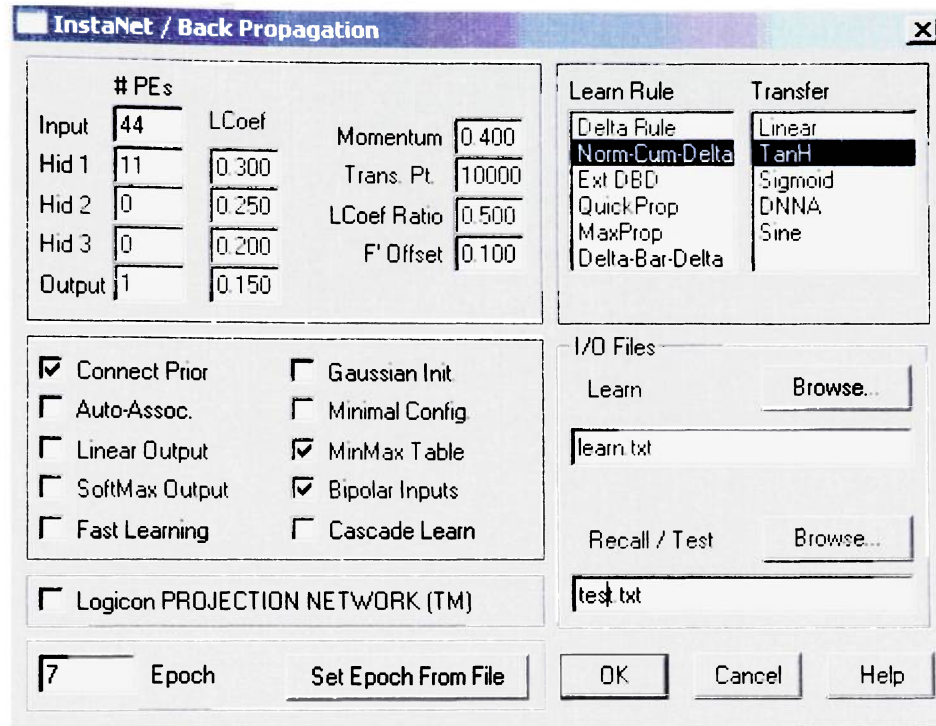
BACKPROPAGATION SETUP:

Figure 46: NeuralWorks setup screen for backpropagation network

Figure 46 shows the setup screen for the backpropagation networks utilized in this research. Fourty-four inputs (one for each amplitude value and 3 categorical variables) were used with an 11 neuron hidden layer (5 expected failure mechanisms) and the output is a single neuron representing burst pressure. Epoch size is determined automatically by the software based on specimen data file size. A normalized cumulative delta learning rule was used with a hyperbolic tangent transfer function. All of the other values

(momentum, transition point, learning coefficient, F' offset, etc) were set to their default values, which had been shown to work well in previous research [Karl, Dion, Spivey, 2006].