UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AUTOMATIC PICKING AND CLASSIFICATION OF ACOUSTIC

EMISSION EVENT ARRIVALS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree

MASTER OF SCIENCE

By

ALVARO A. ORTIZ
Norman, Oklahoma
2010

AUTOMATED CLASSIFICATION AND PICKING OF ACOUSTIC
EMISSION EVENT ARRIVALS

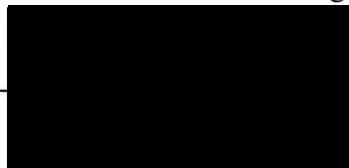A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

_____
Dr. Hong Liu, Chair

_____
Dr. Carl H. Sondergeld

_____
an

ACKNOWLEDGEMENTS

I want to express my most sincere gratitude to Dr. Carl Sondergeld to trust in me and my capabilities to develop this research. I feel very luckily to have him as my advisor and also to introduce me in the geophysics and petroleum industry.

I would like also thank the members of my committee Dr. H. Liu and Dr. J. Fagan for their valuable discussions and thoughts until the final completion of my Master studies.

Special thanks to the faculty and staff of the Mewbourne School of Petroleum and Geological Engineering of the University of Oklahoma and the School of Electrical and Computer Engineering. Big thanks to Gary, Bruce, Shalli, Sonya and Lynn.

To all my family: Mom, Dad and my sister Silvia; without them I would never be here. Anna, this is also for you (for giving me always a smile).

To my beloved and unforgettable friends in Oklahoma and all around the world (Camilo, Pablo, Hernando, David, A. Guerrero, A. Castano and specially Ka to keep my head over the stars), I will always be a Sooner.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xiii

ABSTRACT

Accurate automatic classification and picking of arrival times of events from acoustic emission (AE) signals is of considerable importance for rapid identification and location of seismic events. Due to the large number of digital signals that could be acquired during a simple experiment, manual classification and arrival picking become impractical and subjective.

A system that detects and stores seismic signals generated during hydraulic fracturing (HF) experiments has been used in laboratory experiments to study the mechanism of HF. This system employs a combination of sensors, preamplifiers, signal conditioning unit and a data acquisition (DAQ) module attached to a personal computer (PC). The system captures the AE signals using a simple threshold value.

External and internal noise due to the electronics, sensor coupling, reflections and echoes from the microseismic signals (MS) are captured too. In order to reduce the number of false AE signals, improve the processing time and obtain the most information from the HF processes, an automated solution has been developed to classify, pick the first arrival, and the polarization of AE signals that have been previously captured by the AE monitoring system.

1 INTRODUCTION

Acoustic emissions (AE) signals are noises made when materials deforms or fracture. This is a phenomenon widely used in different disciplines and began to be investigated in the middle of the 20th century. The advances in technology have let the area of AE move forward as an important technique in the study of materials. Ultrasound analysis and non destructive testing (NDT) are just a few of possible techniques that use AE as a tool. Different disciplines like civil engineering, earth sciences, mechanical engineering, metallurgy, etc. have received special benefits from research in AE, and combined with the technological advances in electronics and computer science have led to the develop of new and better ways to detect and analyze the signals.

AE is similar to seismology and can be related to the study of sudden movement of the earth's crust that generates elastic disturbances, known as seismic waves. These waves propagate from the origin spreading spherically in an isotropic material. The waves generated are recorded by seismometers that capture the amplitude vs. time in plots known as seismograms.

AE can be considered also as a form of microseismicity that is generated during the failure process as materials are stressed to failure. It is defined as the

spontaneous release of localized strain energy in stressed materials (Grosse and Ohtsu, 2008). This energy can be recorded by transducers (sensors) placed on the material and analyzed for research to study the mechanism of failure (see Figure 1.1).
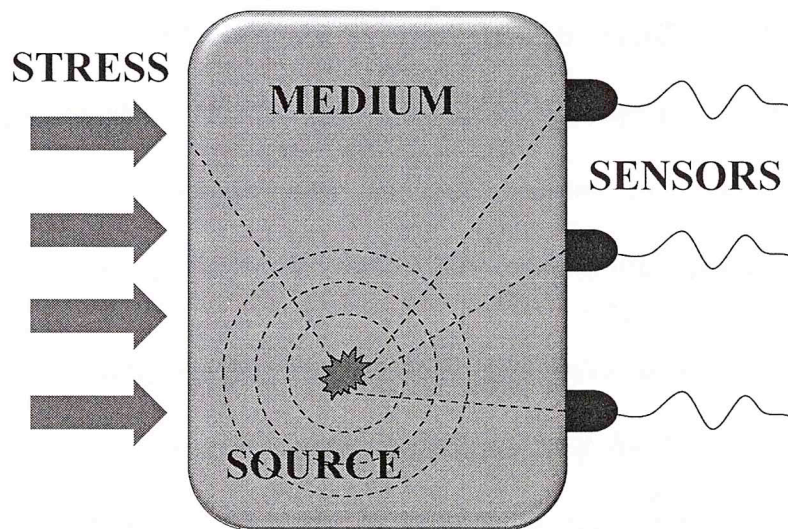
Figure 1.1 Example of a material under stress. The source is represented by a release of energy in red. AE waves (black dashed lines) move in a spherical path reaching the walls of the medium.

The sources of AE signals have widely varying characteristics due to the differences in medium and modes of failure. Monitoring of continuous AE can be used to control the operation of machines, or to locate the origin of earthquakes or defects. Earthquake location has historically depended upon the ability of human analyst to estimate arrival times. During generation of AE two waves are generated, a P-wave and an S-wave. P-wave has an early arrival

which is normally two times that for an S-wave. So first arrivals are P-waves followed later by S-waves. An arrival time of a seismic signal is considered a first break difference between the signal and the noise background.

Generation of AE refers to the source and how it is produced, naturally or induced. Detection and analysis involves human interaction. Most acoustic emission signals are at levels which are outside the normal range of the human ear and for visualization and analysis require the use of electronic instrumentation due to the amplitude and frequency of the signals.

Detection and classification of events typically needs an experienced analyst. This method is known as the manual method. And the person who manually analyzes these signals are known as seismic analyst, this person detects and picks the arrival time of an AE signal. An event is also known, as an AE signal or a group of signals, which possesses certain amplitude and frequency characteristics. The number of events generated depends on the magnitude and rate of stress application and the material tested. The number of recorded signals depends on the number of sensors used to capture the events and the number of signals that are finally captured can be quite large (For an in situ hydraulic fracture experiment, the number of events range from 100's to 1000's).

As mentioned before, localization of the AE events plays an important role in determining the characteristics of a material. AE signals vary considerably in

frequency and amplitude. These signals must be classified to differentiate a true AE signal from noise or a false signal generated from echoes or external noise (see figure 1.2). Following a perfect classification, only true AE signals remain.

The next step is to extract arrival times of first and subsequent waves in these signals. The ability to perform accurate automatic classification and arrival picking on a large number of signals remains a serious challenge facing the microseismic and seismological community. The routine classification and picking is done manually by visual inspection of each signal. For practical applications this is time consuming and demands experienced personal.

The best picking system widely recognized is the human analyst (Alderson, 2004). Manual picking remains as one of the slowest and most repetitive tasks. Analysts are slow, and sometimes they produce questionable results due to the boredom induced by the highly repetitive nature of arrival picking and visual inspection for classification, especially during long periods of intense seismic activity (Alderson, 2004).
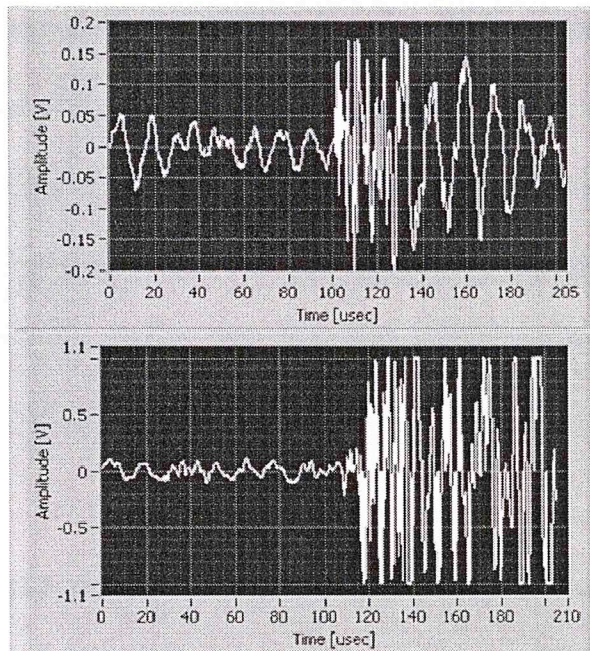
Figure 1.2 Example of AE signals obtained from a hydraulic fracture event. Top: waveform with low signal amplitude but with a distinct arrival. Bottom: signal with a better signal to noise ratio (SNR).

AE events are very simply signals. Signal processing techniques, widely used in areas as electrical engineering, system engineering, and applied mathematics, deals with the operations and analysis of signals (continuous or discrete, time or frequency) to enhance, extract or limit information content.

The advances in areas as digital signal processing (DSP) are applied across many disciplines. Engineering, aerospace, electronic, medicine and earth sciences just to name a few. The world is filled with signals: voltages generated by the heart and the brain, radar and sonar echoes, seismic vibrations and countless other sources. DSP is the tool and the science of using computers to analyze these digitized and discrete signals.

Current AE analysis methods require the user to have an understanding of the relationship between the source and the resulting waveform characteristics, and then analyze the signals by visual inspection to classify and choose the arrival times. When an experiment is carried out, many events occur. Thus, methods that can automatically analyze the data and extract information from these signals, allow timely decisions to be made, and reduce needs for skilled personal. .

Challenges to be addressed by automated classification and detection of arrivals can be summarize as follows:

- The number of events is large; normally around 100's to 1000's or even more occur. If we multiply this number by the number of sensors used (in our experiment 16 sensors are used) we have an extremely large number of signals to analyze.

- An event differs from others by amplitude, frequency and also in the number of sensors capturing the signal. AE events from HF are further shaped by their radiation patterns and path dependent losses due to heterogeneity and attenuation.

- The background noise is different for every signal. This background noise increases the error in manual classification and arrival determination.

- Except for the most impulsive onsets, equally competent analyst will pick onsets at different times.

The main purpose of this work is to develop an automated way to classify and pick the first wave arrivals (P-wave signals) from large volumes of digital microseismic signals. The foundation for this investigation is in exploiting the amplitude and frequencies of the recorded signals and developing an algorithm for classification.

1.1 Motivation

A strong motivation for accurate AE event detection and location of seismic activities is to duplicate or even replace the performance of an expert operator or seismic analyst. Since the beginnings of AE and its acquisition, the problem of the data volume, false events and noise has limited studies, application and analysis of the results.

Hydraulic fracture is a common technique that improves well production and extraction of oil and gas, especially in tight formations like Barnett shale (Castano, 2010) and it is critical to the future development of shale gas and tight gas resources for the U.S. and the world. The hydraulic fracture processes and the nature of associated AE events are complex. This makes automatic processing challenging. This thesis work will try to improve the process of classification and arrival picking of AE events generated during hydraulic

fracture experimentation in laboratory scale. The results obtained by the automated process will be compared to the results obtained by the seismic analyst.

1.2 Research Objectives

The principal objective for this research can be summarized:

Generate an automated computer algorithm that can duplicate the performances of a human operator to classify and pick the arrival time of AE signals generated by HF process.

Other objectives or sub-objectives of this research are:

a. Generate results in a form compatible with other post-processing programs.

b. Analyze AE associated with hydraulic fractures to determine AE sources (location), focal mechanism and source parameters.

c. Initial classification to eliminate false events and save only the events and signals that meet certain user characteristics.

d. Classification of "good events" and "bad events" in order to analyze the best events with the automated arrival picking method choose.

1.3 Acoustic Emission Signals

It is generally accepted that any kind of material generates a sound or tick when is stressed to failure; some sounds are generated by the sudden change in the crystal structure. AE waves in general, are elastic waves in a solid. The frequency range of the radiated energy covers a wide frequency spectrum, ranging from audible to ultrasonic (Grosee and Ohtsu, 2008). In laboratory experiments we detect ultrasonic waves.

A variety of terms, including AE, microseismic activity, seism-acoustic activity, subaudible noise, roof and rock talks, elastic shocks, elastic radiation and micro-earthquake activity are utilized by various disciplines to denote this phenomenon (Hardy, 2003). Throughout this thesis, this phenomenon will be referred as Acoustic Emission/ Microseismic (AE/MS) activity.

This thesis is focused on the techniques used in the geotechnical areas. In geologic materials the origin of AE/MS activity appears to relate to the process of deformation and failure which are accompanied by a sudden release of strain energy. AE/MS activity may originate at the micro-level as a result of dislocations and at the macro-level by twinning, grain boundary movement, or initiation and propagation of fractures through and between mineral grains, in the mega-level it will be found by fracturing and failure of large areas of material or relative motion between structural units (Hardy, 2003).

Lavrov (2005) also defines an acoustic emission as a phenomenon of emitting elastic waves as a result of irreversible or partially reversible changes in the structure of a solid under the action of various external and internal physical factors.

Muravin (2008) defines an AE emission as a phenomenon of sound and ultrasound wave radiation in materials undergoing deformation and fracture processes.

1.4   Arrival Picking Theory

The true onset time of a seismic signal could be defined as the moment when the first energy of the particular signal arrives at a sensor. Every signal is normally identified by a difference in amplitude and frequency content, or even wave polarization from the background noise.

The onset time is usually picked as the point where the difference from the background noise first occurs (Leonard, 2000), (see Figure 1.3). Normally an experienced analyst is able to pick an arrival that is hidden in noise based on experience (Leonard, 2000).

Methods for reducing the noise without distorting the signal will normally result in an earlier arrival time being picked (Leonard, 2000). Filtering or noise

reducing techniques need to be applied with caution so they do not eliminate the

real arrival time.
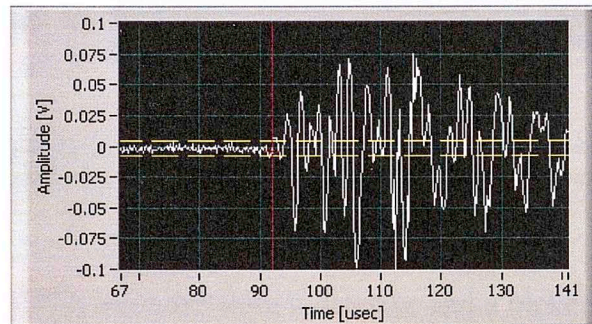
Figure 1.3 Example of a HF signal. Red line show the arrival acquired by an automated method. The dashed yellow lines show the amplitude of the background noise.

1.4.1   Automatic Time Pickers

An automatic arrival time picker is a system able to emulate the behavior of an

experienced analyst in the picking of the onset time from an AE arrival. The

process is shown schematically in figure 1.4 as an input - output functional. The

input will be the signal or group of signals previously selected as a real seismic

event and the output is the arrival time picked for that group of signals.

Figure 1.4 Input - output block diagram representation of the automatic picking process to extract the arrival time of a signal or group of signals.

It is important to differentiate between a detector and a picker. A detector is able to recognize an AE event over the background noise without picking the arrival. In other words a detector scans an AE signal and determines if it is an event that meets the characteristics set by the detector and if the quality is sufficient for storage, while a picker performs more precise analysis required for hypocenter location and further study. The most important difference between a time picker and a phase detector is the required precision of timing of the onset of first arrival (Allen, 1982).

It is important to mention that there is not a single method that ensures consistent onset time picking for every arrival. Any particular method will fail when the difference between the noise and the signal is small, particularly when the Signal to Noise ratio is low (Bai and Kennet, 2000). With this understanding, it is necessary to analyze which automatic method will be the most robust and satisfactory.

Bandpass filtering will not work when the noise and the signal have almost the same frequency content (Bai and Kennet, 2000). This occurs due to the reflections of the original waves between interfaces and discontinuities on the medium. Most significant reflectors are the walls of the sample that reflect and send back the signal as an echo with frequencies close to the fundamental.

1.5 Hydraulic Fracturing

Hydraulic fracturing is a tool to enhance the extraction and production of an oilfield, is commonly known as a stimulation technique that is applied to increase the production rate and enhance hydrocarbon recovery. In the laboratory, these scaled experiments are used to understand the field techniques.

1.6 Background History on AE

According to Scott (1991) and Muravin (2008), the first documented observations of AE may have been made on the 8th century by an Arabian alchemist, Geber, who describes the "harsh sound or crashing noise" emitted from tin. He also describes iron as "sounding much" during forging. But probably the first practical use of AE was by pottery makers, thousands of years before recorded history, to assess the quality of their products (Muravin, 2008).

The successful application of AE in different areas like detection and location of faults on pressure vessels, damage in composites, monitoring of civil

engineering structures (e.g bridges, reactors, platforms, pipelines, etc.) and in general in the area of geotechnical engineering has been creating different branches and rapidly advancing new techniques.

The first experiments using AE were related to the measurement of the mechanical stability of rock materials and the associated rock structure in the mines and tunnels. In the late 1928, A.F. Ioffe published a paper, "the physics of crystals," which can be considered one of the first researches on AE in rocks (Lavrov, 2005; Scott, 1991, and Muravin, 2008). Ioffes's paper mentioned that each particular event on the deformation of rock salt is accompanied by a "noise" and indicates the possibility of using this noise for studying the nature and behavior of the deformation. During the 30's, Foster and Scheil (1936) discussed the clicks which occurs during the formation of martensite in high-nickel steel measuring the small voltage and resistance variations caused by sudden transformations in this materials Scott, (1991) and Muravin, (2008).

By 1941, two researchers of the U.S Bureau of Mines (USBM), Obert and Duvall, discovered that a stressed rock pillar appeared to emit micro level sounds. This discovery was observed in a laboratory and called in nontechnical terms as "rock talks" (Hardy, 2003; Scott, 1991). Mason, McSkimin and Schockley, (1948), suggested measuring AE to observe the moving dislocations

by means of stress waves they generated, on ultrasonic of twinning in tin (Scott, 1991; Muravin, 2008).

Later, on the 50's, AE was recognize as one of the best ways to monitoring rock fractures generated in stressed rocks in coal mines (Lavrov, 2005). In 1950, Josef Kaiser (Germany) used tensile tests to determine the characteristics of AE in engineering materials. The result from his investigation was the observation of the irreversibility phenomenon that now bears his name, the Kaiser Effect, (Hardy, 2003; Scott, 1991; Muravin, 2008). It is generally known that his research represents the beginnings of AE as practiced today (Scott, 1991). The first extensive research after Kaiser was done in the United States by Schofield in 1954. Schofield investigated the application of AE in the field of materials engineering. He concluded that AE is mainly a volume effect and not a surface effect (Scott, 1991; Muravin, 2008).

During the 60's, according to Muravin, (2008), AE start to be used for true industrial applications. The first test in the USA was conducted by the aerospace industry to verify the integrity of the Polaris rocket motor for the U.S. Navy. During the following years, 1963-65 Dunegan, suggested the use of AE for examining high pressure vessels and founded the first company that specializes in the production of AE equipment (Muravin, 2008).

Due to the use of different terms, i.e., acoustic emission or microseismic activity to denote the general phenomenon of the analysis of noise created under stress in different materials, the research and the associated technical publications in the field have been widely scattered throughout various areas of engineering and science. Around the 70's and 80's as a result of the continuing efforts of Acoustic Emission Working Groups in the USA (AEWG), Europe and Japan; different Symposia sponsored by the American Society for Testing and Materials (ASTM), AEWG, and the Society for Nondestructive Testing (ASNT) joined together to integrated the different fields into a common language and share all the information related; workers in different disciplines are becoming better acquainted with the extensive literature available for the subject (Hardy, 2003). Today, AE is a commonly method used in different fields besides geotechnics and petroleum engineering.

1.7   Synthetic Signals

In order to determine how effective the automatic picking algorithm is, it is necessary to generate synthetic signals that emulate or try to emulate the behavior of AE signals.

The principal characteristics of synthetic signals could be summarized by:

- Known arrival time location
- Controllable amplitude

- Controllable frequency: singular and multi-tone.

- Repeatable

A test has been prepared to examine the behavior of the automatic picking algorithm. One waveform recorded as AE wave is equivalent to the sum of sinusoidal waves with a principal fundamental frequency. Our synthetic AE signal $S(t)$ consist of an exponentially decaying sine wave with a dominant frequency, determined statically on several samples of real AE signals using the analysis of the fundamental frequency peak in the Fast Fourier Transform (FFT).

$$S(t) = A_p \exp(-\alpha t) \sin(wt)$$

Where A_p correspond to the amplitude of the sine wave with w frequency in Hz, which is modulated by the exponential decaying curve by the factor α. The signal's S(t) frequency, w, is determined to be 350 kHz. The synthetic signal is illustrated in the figure 1.5, where the arrival time has been selected at $t = 60\ \mu sec$.
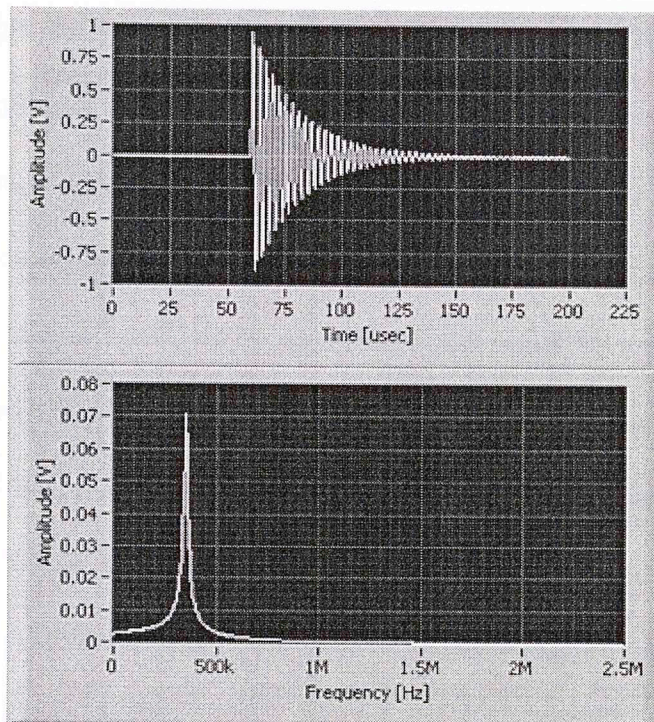
Figure 1.5 Synthetic signal S(t). TOP: time signal with a clear arrival time located at 60 μsec. BOTTOM: spectral bandwidth representation of signal. ZOOM: spectral bandwidth around 350 kHz.

Real AE signals are observed mixed with noise. Sources of noise are produced by electronic equipment and echoes produced by the wave reflected between the impedance contrast inside and outside the rock. The noise is distributed in and around the bandwidth of the true signals. These two types of noise are simulated. Background noise which is normally produced by electronic equipment can be reproduced by Gaussian White Noise with variable sampling ratio, number of samples and standard deviation. The other noise attributed to echoes and reflection of the wave inside the rock can be simulated by a multi-tone sine wave. This multi-tone signal will be represented by a group of

18

frequencies, also known as tones, between 10 kHz and 500 kHz. Figure 1.5 shows an example of these two types of noise; Gaussian white and multi-tone. The FFT can be used to measure the RMS (root mean square) amplitude of the multiple frequencies and noise components of a digitized signal. This example of noise clearly shows how complicated the extraction of the arrival information can be. The arrival picking due to noise presented on the signals can be interpreted in different positions of time by different seismic analyst. The same is applicable to an automatic pick method; the arrival time could be corrupted by the noise presented on the signal. The amount of noise can be represented by the Signal to Noise Ratio (SNR).

SNR is a measure that represents the difference between the level of signal and the level of noise, and is normally represented in dB (decibels) following the equation 1.1

$$SNR = 20 \log \left(\frac{S}{N}\right) [dB]$$

Where, S represents the signal and N the noise or unwanted signal. SNR can be represented in different way and the definition has been analyzed and studied in different areas that require the handle signals.
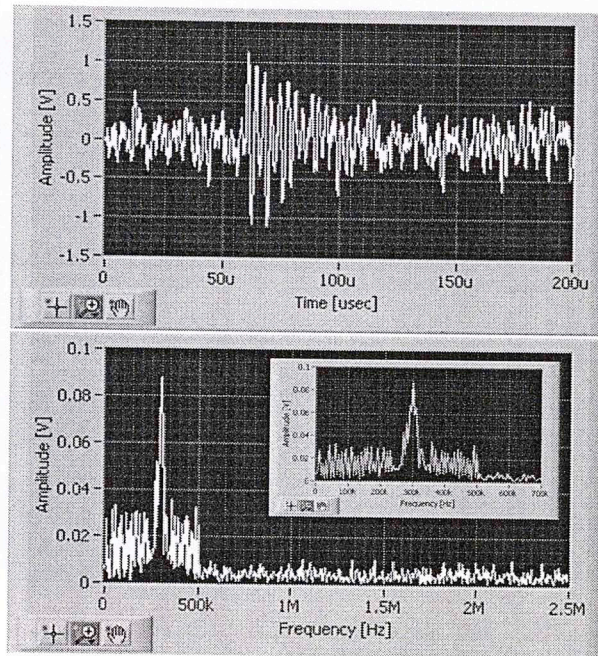
Figure 1.6. Synthetic signal with noise added. TOP: time signal covered by Random Gaussian White Noise with a standard deviation of ± 150 mV. BOTTOM: Spectral frequency content of synthetic signal plus Gaussian and multi-tone noise.

I systematically varied the SNR to test the robustness of the picking method. Based on these tests, we can determine at what SNR the picking will fail. Figure 1.7 shows the results of the signal with different SNR values.
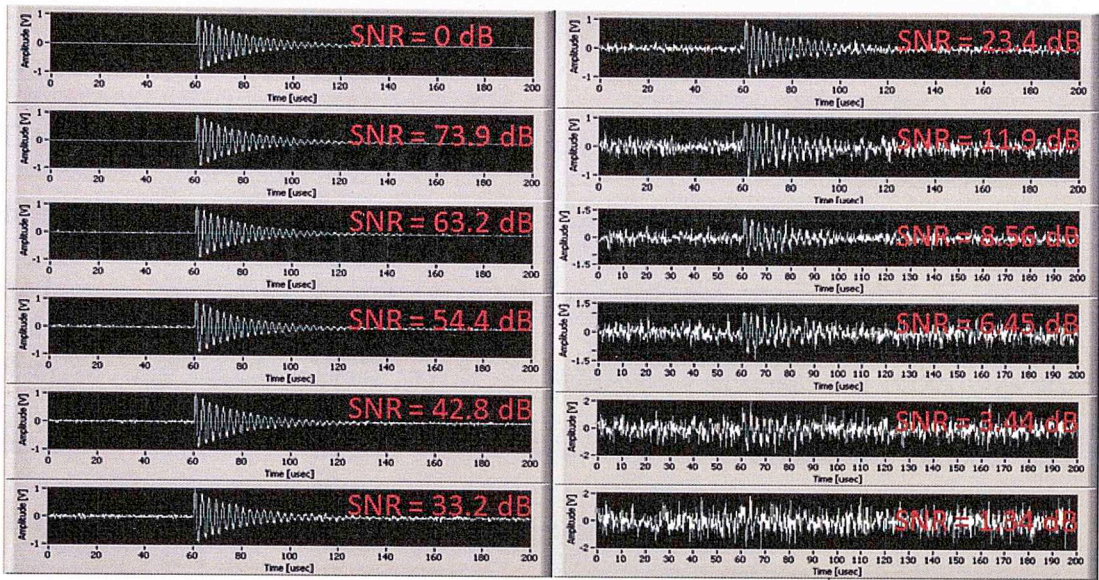
Figure 1.7 SNR variation using multi-tone signals and random Gaussian white noise added to original signal.

2   CLASSIFICATION OF EVENTS

The current manual event classification has been recognized as a repetitive and difficult task. The repetition of a manual task brings numerous errors induced by the human mind acting in a situation of stress. Misclassification of large number of events and detecting good events in the presence of noise and reflections are the most commons problems encountered. This time consuming process and verification of errors becomes a costly task and limits the number of useful events.

In the seismic monitoring industry they commonly implement a simple amplitude threshold for event classification. A threshold value in amplitude is a common and easy method to differentiate the background noise (in the order of mV) from the amplified microseismic signal. Following a simple threshold classification method will capture reflections as "good" events. These reflections are in the same order of amplitude as a real signal and are able to change and complete mask the arrival of a seismic and the secondary S waves.

Observation of large numbers of microseismic signals (specifically "good" events) shows that generally these signals have lower dominant frequency content, shorter P-wave event lengths and flatter time domain characteristics (Tan, et al., 2007). Based on these observations a classification algorithm can be developed involving different statistical measures. Statistical analysis provides a

a strong technique for microseismic signal classification which is superior to common industry techniques such as thresholding, amplitude analysis, frequency filtering and event-length detection. For this reason, an extensive testing was performed to determine which classification algorithms involving statistical analysis and other characteristics were most robust and accurate.

The resulting characteristics are a multidimensional data set. Multivariate data reduction techniques such as principal component analysis (PCA) are able to reduce the multidimensionality of these data and simplify the classification.

The purpose in this chapter is to analyze and combine microseismic signal analysis algorithms to develop a precise automatic classification method for a large numbers of AE signals. The system must be able to differentiate "good" signals and "bad" signals obtained from HF testing on rock samples.

2.1 "Good" signals vs. "Bad" signals.

After reviewing several HF experiments, certain characteristics differentiate a good signal from a bad signal. These characteristics are generally visible and easily extracted from the raw data.

Good signals normally present the following characteristics (Tan et al., 2007):

- Lower signal variance. Variance is a measure of statistical dispersion. It indicates how values vary about the mean. Larger dataset variances

correspond to larger expected deviations of arbitrary data points from the mean (Tan, 2007); which means that the signal is covered by noise acting randomly and dispersing the values.

- Higher central data distribution. The distribution around a central point is another measure of randomness in the signal. This is measured by counting the number of points lying outside a mean centered window (Tan, 2007)

- Less frequent oscillations. It is observed that the magnitudes of signed amplitude differences between adjacent time series data points were generally greater for noise microseismic traces compared to "good" traces (Tan, 2007).

These characteristics added to an analysis of signal to noise ratio (SNR), frequency and amplitude analysis and proper analysis of the pre-trigger signal and the arrival of the signal (acquisition process) will result in the right combination of characteristics for the classification process.

Manually extracting these characteristics from a signal is both difficult and subjective. The human brain can interpret a noisy signal as a good one just because at first glance (no zoom applied) an arrival could be extracted. For a good analyst three levels can be easily distinguished in the classification of signals:

Good signals: like the one presented in figure 2.1, (SNR around 8 to 40 dB) which have clear arrivals, low background noise and no reflections.

Medium signals: figure 2.2, (SNR 8 to 2-3 dB), signals often classified by a non-expert analyst as "good" ones but a manual arrival picking is distorted by the noise and normally incorrect.

Bad signals: (see figure 2.3), are signals completely obscured by noise and reflections, with no possibility to extract an arrival or any other useful information.

An automatic method to classify these signals needs to not only differentiate clearly a good signal from a bad signal but also classify medium level signals that have an unclear arrival times and cannot be compared by manual methods.
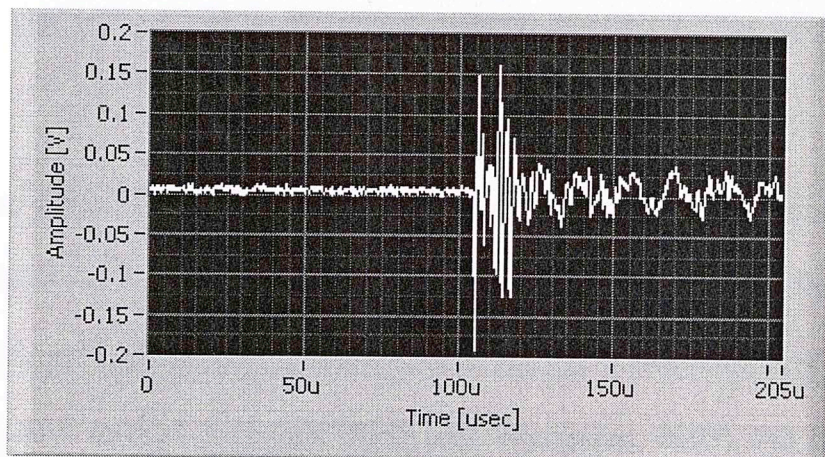
Figure 2.1 Example of a "good" signal
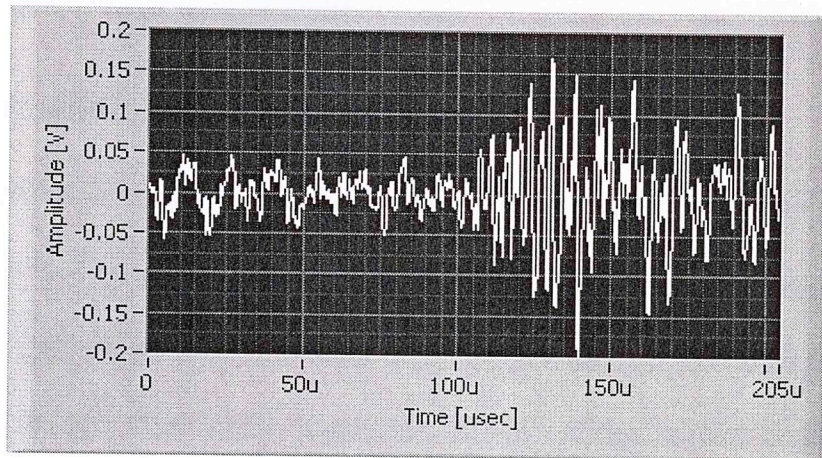
Figure 2.2 Example of a "medium" signal
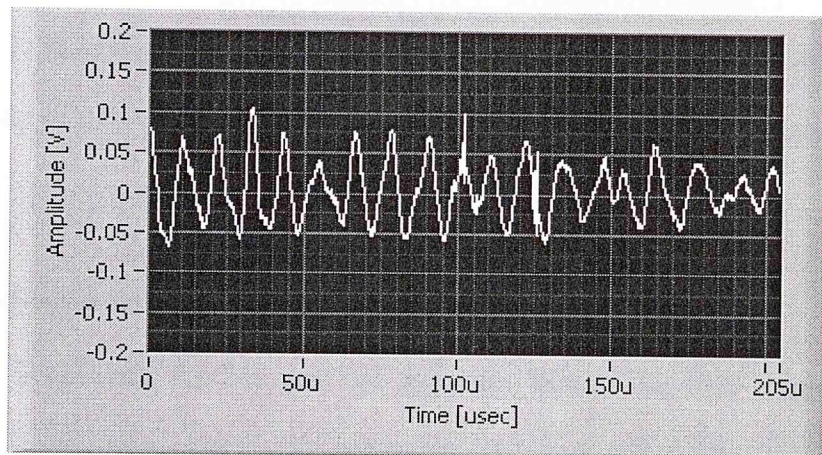
Figure 2.3 Example of a "bad" or "noise" signal

In order to classify an event, a certain number of sensors must record "good" signals and be classified as "good". One of the final objectives of the HF data processed is to locate the hypocenter of the event. For a 3-dimensional location a minimum of 4 sensors is needed (Stein and Wysession, 2003), but in order to minimize the amount of uncertainty and error, more sensors are required. In

laboratory experiments, the minimum number of sensors needed to classify an event as "good" is defined to be six.

2.2 Classification Algorithms

Several criteria and algorithms have been used previously to classify signals. According to Tan (2007), the best results for classification are obtained using statistical approaches in combination with a proper method to reduce the amount of data obtained from every algorithm. Based on the normal characteristics found for good signals, seven different characteristics algorithms were designed to classify events. Combinations of these characteristics are finally used to classify the events along with the multidimensional data space reduction using principal component analysis (PCA).

2.2.1 Signal to Noise Ratio (SNR)

Signal is considered as the part of the data where we can extract all the necessary information. On the other hand, noise is all the unwanted information which also occur this portion of the data. A clear definition of signal and noise is subjective in many cases for the people who know how to interpret the data, for others, everything could be noise.

In the acquisition procedure for HF signals (chapter 4) a configuration for pre-trigger setting in the acquisition software, gives us a beginning for the identification and differentiation between signal and noise. Signal pre-trigger

points determine the amount of memory to be included in the data capture prior to the trigger time. A trigger time is the moment in which a signal on any given channel crosses the threshold value (0.01 V).

Figure 2.4 shows the acquisition of signals using a different percentage of pre-trigger points (10%, 25% and 50%) on pencil-break calibration test. These different configurations do not determine the exact arrival time, just the point around which the arrival occurs. . Figures 2.1, 2.2 and 2.3 show an example of signals using 50% pre-trigger points on a total of 1024 data points, this means 512 data points or 102.4 μsec of data before threshold value is reached. It is easy to identify that a "good" signal should behave as the one presented on figure 2.1 and 2.4.

The SNR is normally measured in the frequency domain by analysis of energy of the signals around the fundamental frequency. In the time domain, a special arrangement can be done taking the advantage that the number of pre and post signal arrival can be approximate from the pre-trigger points. To accomplish this analysis, a window after and before the pre-trigger point is taken and assuming that the window before contains noise and the window after contains the real signal.
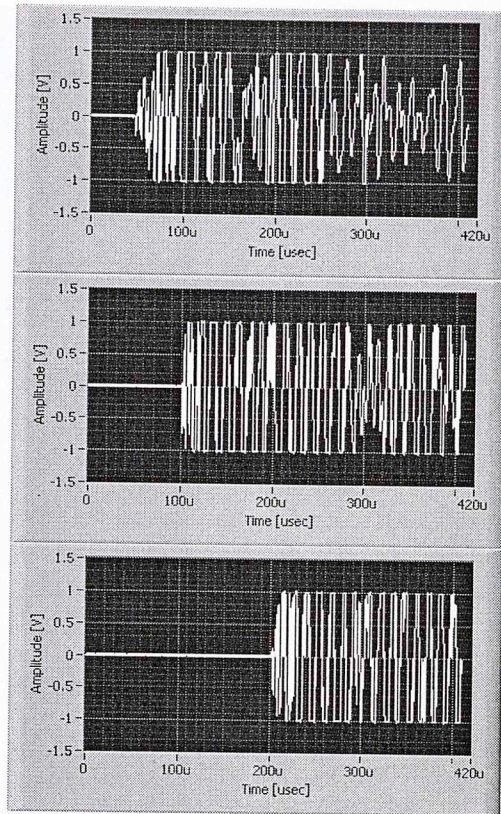
Figure 2.4 Pencilbreak test using 10%, 25% and 50% (top to bottom) pre-trigger points, note that the total acquisition trace length is 2048 points

Figure 2.5 shows an example of the window used in analyzing SNR during a

lead-break calibration experiment (top), and a real HF signal (bottom).
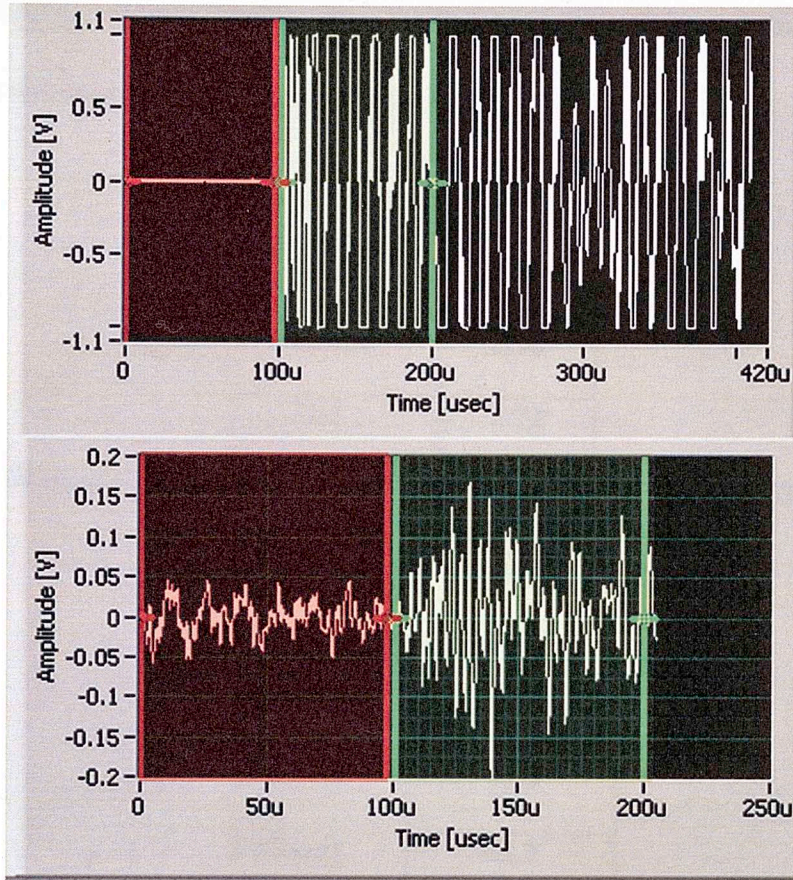
Figure 2.5 Windows for analysis of SNR. Red zone will measure unwanted noise signal. Green zone will measure the real signal obtained from an AE experiment. The windows are set knowing the pre-trigger points setting. Top figure represent a 25% pre-trigger point. Bottom signal use 50% of pre-trigger points

The following algorithm defines the value of SNR extracted from every signal:

$$SNR = 20log_{10}\left(\frac{A_{signal}}{A_{noise}}\right) = 20log_{10}\left(\frac{\sqrt{\sum_{i=ws}^{ww} X_i^2}}{\sqrt{\sum_{i=0}^{ws} X_i^2}}\right)$$

Where, A represents the amplitude of the signals, ws the window for noise portion of signals (defined by pre-trigger) and ww the window for signal (same

number of points of *ws*). Results from SNR measurements in real and synthetic

signals are summarized in the following tables:

Event Reference	SNR (dB)
Good	30.84
Medium	2.90
Bad	2.72

Table 2.1 SNR measurements on real signals figures 2.1, 2.2 and 2.3 showing different value levels on dB

Event Reference	SNR (dB)
1 (Good)	(Inf) No Noise
2 (Good)	73.9
3 (Good)	63.2
4 (Good)	54.4
5 (Good)	42.8
6 (Good)	33.24
7(Medium)	23.44
8 (Medium)	11.9
9 (Noise)	8.56
10 (Noise)	6.45
11 (Noise)	3.44
12 (Noise)	1.34

Table 2.2 SNR measurements on synthetic signals using different levels of background noise

2.2.2 Threshold Percentage Algorithm

Visual inspection and comparison of "good" and "bad" events provides several conclusions. Applying statistical analysis on time sequence plots shows that time is an important factor that contributes to variability of data. A time series or time sequence is a data set in which the observations are recorded in the order in which they occur (Montgomery, et al, 2001). A seismogram time series plot is a graph in which the vertical axis denotes the observed value of the variable (in our case, the amplitude in V), and the horizontal axis denotes the time (μsec).

A pre-trigger signal is set during the acquisition that provides a time window for analysis of background noise. If the pre-trigger is set properly, the window can enhance the picking of the arrival time. Figure 2.1 indicates that this pre-trigger position is set at 100 μsec. This leaves enough time to analyze, differentiate and find the proper time difference where the signal arrives.

Following Tan (2007), he concludes: "more noise data points would be expected to lie outside an arbitrary mean-centered amplitude window w than good data points." To validate Tan's conclusion, a defined window has been created around the mean, which is zero after DC removal from the signals. The threshold algorithm simply counts the number of data points lying outside the threshold limits $-a \leq w \leq a$. Table 2.1 and 2.2 summarize the percentage of outlying data points using a threshold value of $a = \pm 0.01$.

Event Reference	SNR (dB)	% Points Outside
Good	30.84	33.3
Medium	2.90	36.2
Bad	2.72	73.8

Table 2.3 Percentage of points lying outside a window threshold value of 0.01 on real example signals from HF (see figures 2.1, 2.2 and 2.3)

Event Reference	SNR (dB)	% Points Outside
1 (Good)	(Inf) No Noise	38.1
2 (Good)	73.9	40.5
3 (Good)	63.2	48.5
4 (Good)	54.4	62.9
5 (Good)	42.8	76.0
6 (Good)	33.24	83.2
7(Medium)	23.44	90.5
8 (Medium)	11.9	95.2
9 (Noise)	8.56	93.8
10 (Noise)	6.45	95.7
11 (Noise)	3.44	98.6
12 (Noise)	1.34	98.9

Table 2.4 Percentage of points lying outside a threshold window of 0.01 on synthetic signals with different SNR ratios

Tan's conclusion is derived from the Chebyshev's inequality, this analyzes the likelihood of dataset points existing a given distance away from its mean. A large data variance will correspond to an increase of data points lying outside a centered window $2a$, according to the following equation:

$$\Pr\left(|X - E[X]| \geq a\right) \leq \frac{VAR[X]}{a^2}$$

In this equation, $\Pr\left(|X - E[X]| \geq a\right)$, is the probability that a random variable X will lie, at least, outside a distance a away from the mean $E[X]$.

Analyzing the results of tables 2.1 and 2.2, the signals with a high SNR (generally known as "good" signals) contains fewer outlying data points.

2.2.3 Zero Crossing's Algorithm

According to Tan, (2007): "microseismic noise signals tend to oscillate more frequently about the time axis and that magnitudes of signed amplitude differences between adjacent time series data points are generally greater compared to good traces." These conclusions are also derived by visual and statistical interpretations.

To evaluate Tan's conclusions, an algorithm that counts the number of zero crossings over a defined window in a signed amplitude range $-z \leq v \leq z$ was

developed. The algorithm first removes all the data-points that fall within the defined window in order to removes low-amplitude noise (Tan, 2007).

The amplitude of adjacent data-points on noise signals generally vary on greater than good signals (Tan, 2007). So the step that eliminates the low-amplitude data points is trying to preserve sign changes but eliminate this change on "good" signals to enhance the difference between "good" and noise signals.

Zero-crossing algorithm generates a TRUE immediately after the transition occurs on any direction (minus to plus or plus to minus) and finally counts the total number of "true" and generates a percentage of zero crossings over the total number of points. This algorithm was tested on real and synthetic signals. Table 2.3 and 2.4 shows the result on the percentage of zero-crossing with low-amplitude noise elimination.

Event Reference	SNR (dB)	% Points Outside (Window = 0.01)	% Points Outside (Window = 0.03)
Good	30.84	3.32	1.56
Medium	2.90	7.62	4.49
Bad	2.72	3.32	1.95

Table 2.5 Percentage of zero crossings count on real signals using 2 different windows of 0.01 and 0.03V

Event Reference	SNR (dB)	% Points Outside (Window = 0.01)	% Points Outside (Window = 0.03)
1 (Good)	(Inf) No Noise	6.4	4.8
2 (Good)	73.9	7.4	4.8
3 (Good)	63.2	11	5.2
4 (Good)	54.4	17.2	5.8
5 (Good)	42.8	24.6	13
6 (Good)	33.24	31.6	21
7 (Medium)	23.44	35.6	27.8
8 (Medium)	11.9	41.2	37.6
9 (Noise)	8.56	44.6	40.6
10 (Noise)	6.45	44.2	41.2
11 (Noise)	3.44	48.4	46.2
12 (Noise)	1.34	49.4	48

Table 2.6 Percentage of zero crossings count on synthetic signals using to different values of windows 0.01 and 0.03 for low-amplitude noise reduction

Tabulated results show a clear incremental changes in the percentage of zero crossings (using a window for low amplitude noise reduction) when the SNR is reduced. Note that noise events generally have a significant number of crossings to zero in comparison with a good signal. This indicates that zero crossing algorithms are able to quantify and differentiate the signals with different SNR.

2.2.4 Histogram Algorithm

A histogram is a graphical representation of a frequency distribution. This distribution is a compact summarization of the data divided in different intervals or bins. A good differentiation between "good" and "bad" signals is that a "bad" signal tends to be distributed closely around the time axis; this analysis can be measured by a histogram.

Using a histogram for data distribution, it is possible to represent the number of signal data points that fall within disjointed amplitude ranges (Tan, 2007). To illustrate how a histogram can represent a signal distribution, it is necessary to divide the range of the data into intervals or bins. The number of bins depends on the quantity of data points and some judgment must be used to select an appropriate number. According to Tan (2007), the simplest way to determine the number of histogram bins is using the following algorithm:

$$n = \left\lceil \frac{max - min}{b} \right\rceil$$

Where n is the total number of bins; max and min represents the largest positive value and the largest negative value, respectively and b represents the width of every bin. An important issue when creating a histogram is to determine the optimal width, b, or analyzing the equation, the appropriate number of bins.

Tan (2007) tried to analyze and discuss three well known equations for an optimal choose of *b*: Sturges (1926), Scott (1979) and Freedman and Diaconis (1981). As a general conclusion, the best solution to determine this number is to do it empirically. No mathematical relationship concerning optimal histogram bin width has been developed for all types of datasets.

The histogram parameters for empirically determined the number of bins are selected from every signal. For example, using as reference the dataset of the figure 2.1, 2.2 and 2.3 that corresponds to a "good", "medium" and "bad" signals and using a number of 99 bins, the central distribution diagram or histogram with DC eliminated can be analyzed on the following figures:
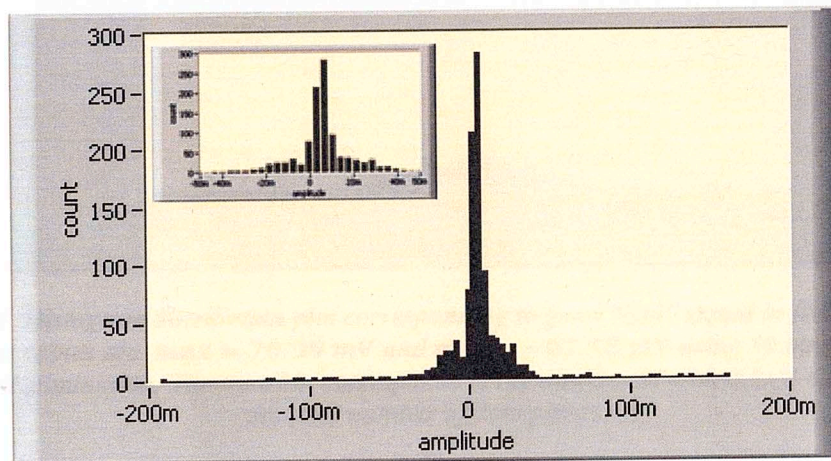
Figure 2.6 Histogram distribution plot corresponding to "good" signal in figure 2.1. The extreme values are $max = 162.45 \, mV$ *and* $min = -192.31 \, mV$ *using 99 bins and DC component elimination. The max bin corresponds to the number 55 with 27% data points of the total number of data points.*
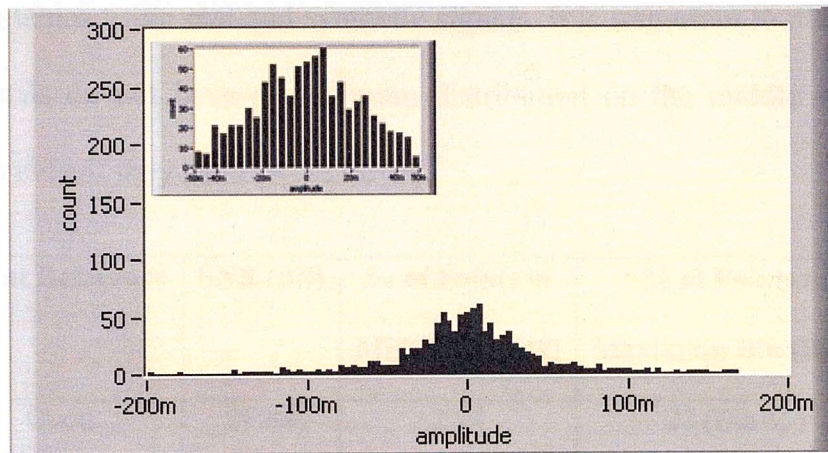
Figure 2.7 Histogram distribution plot corresponding to "medium" signal in figure 2.2. The extreme values are $max = 169.75\ mV$ and $min = -199.58\ mV$ using 99 bins and DC component elimination. The max bin corresponds to the number 55 with 5.86% data points of the total number of data points.
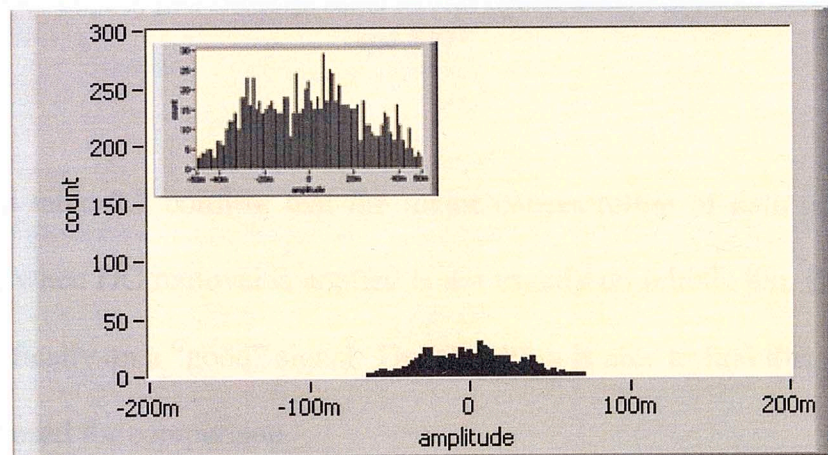
Figure 2.8 Histogram distribution plot corresponding to good "bad" signal in figure 2.1. The extreme values are $max = 70.99\ mV$ and $min = -63.48\ mV$ using 99 bins and DC component elimination. The max bin corresponds to the number 51 with 2.83% data points of the total number of data points.

A good signal is easily identified in a histogram plot because present a concentration of data points on the middle bin or closer. A case using the middle bin and other using the maximum bin component will be analyzed on the

following tables using real and synthetic signals. It is important to mention that some signals do not have a maximum distribution on the middle bin of the histogram even if they have DC removed.

Event Reference	SNR (dB)	% of Points in Middle Bin (49)	% of Points in Maximum Bin (Bin #)
Good	30.84	2.24	27 (Bin 55)
Medium	2.90	5.08	5.86 (Bin 55)
Bad	2.72	1.76	2.83 (Bin 51)

Table 2.7 Percentage of zero crossings count on real signals using 2 different windows of 0.01 and 0.03V

Results on table 2.7 confirm that the major concentration of data points on a histogram when DC removal is applied is not exactly on middle bin. Bin 55 was used specifically on a "good" signal. The algorithm is able to find this maximum bin and is used for comparison.

Event Reference	SNR (dB)	% of Points in Middle Bin (49)	% of Points in Maximum Bin (Bin #)
1 (Good)	(Inf) No Noise	3.4	49.5
2 (Good)	73.9	3.4	36.9
3 (Good)	63.2	7.7	47.0

4 (Good)	54.4	8.1	29.0
5 (Good)	42.8	8.5	22.5
6 (Good)	33.24	8.2	14.3
7(Medium)	23.44	7.0	9
8 (Medium)	11.9	3.7	4.6
9 (Noise)	8.56	3.5	5.3
10 (Noise)	6.45	3.1	5.1
11 (Noise)	3.44	3.24	3.8
12 (Noise)	1.34	3.1	3.2

Table 2.8 Percentage of zero crossings count on synthetic signals using to different values of windows 0.01 and 0.03 for low-amplitude noise reduction

Due to the random nature of the real and synthetic signals, the middle bin (bin 49) does not contain the most of the data points as commented by Tan (2007). In our case and analyzing the results obtained on tables 2.7 and 2.8, an adjustment has been made to analyze the bin that contains or concentrates the majority of the data points.

The middle bin does not show a normal tendency in the percentage of data points due to the variation of the SNR compared to those analyzed by Tan (2007). In other words, when the maximum bin data points are analyzed, the tendency shows that this percentage will reduce with the increasing of the SNR which will help differentiate between a "good" signal and a "bad" signal.

41

2.2.5   Frequency Analysis

AE signals on geologic materials have also been described in terms of their frequency spectra. In general, any transient signal can be considered as a superposition of a large number of sinusoidal signals of specific frequencies and amplitudes. The conversion between time domain (amplitude vs. time) and frequency domain (amplitude vs. frequency) can be expressed mathematically by the Fourier integral, following the general form:

$$S(t) = \frac{1}{\pi} \int_0^\infty S(\omega) cos[\omega t + \emptyset(\omega)] d\omega$$

Where $\omega = 2\pi f$; $S(t)$ represent the amplitude of the wave in the time domain; $S(\omega)$ represents the amplitude of the wave in the frequency domain; f is the frequency; t is the time; and $\emptyset(\omega)$ is the phase factor.

For a continuous signal, frequency domain can be determined using the Fourier transform, which can be defined as

$$S(f) = \int_{-\infty}^\infty s(t) \exp(-j2\pi f t) dt$$

Figure 2.9 shows the signal corresponding to a "good" event and its associated frequency spectra. The frequency spectra show components that reach 2.5 MHz that corresponds to the Nyquist cutoff frequency; however, the major frequency components are in the range of 200 - 500 kHz.

It is possible to compare the frequency content of different signals obtained from an HF experiment. Figure 2.10 and 2.11 shows the frequency content of a "medium" signal and a "bad" signal respectively.
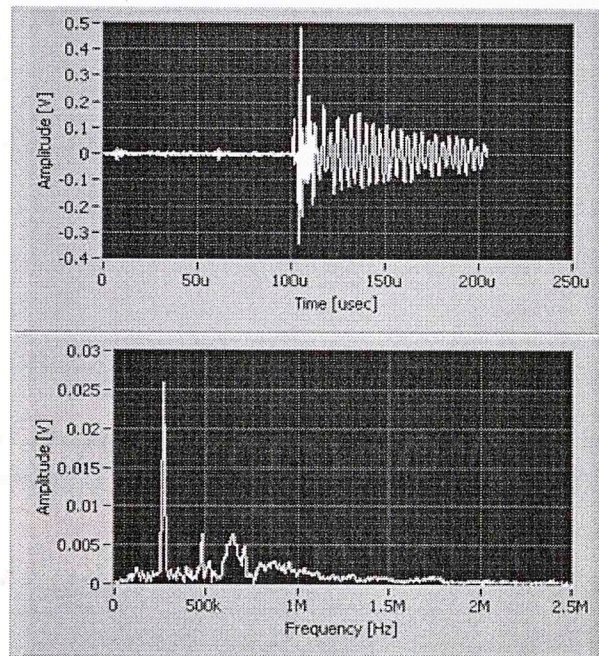
Figure 2.9 TOP: A "good" signal from an AE event in a HF experiment in the time domain. BOTTOM: Frequency spectra using RMS magnitude.
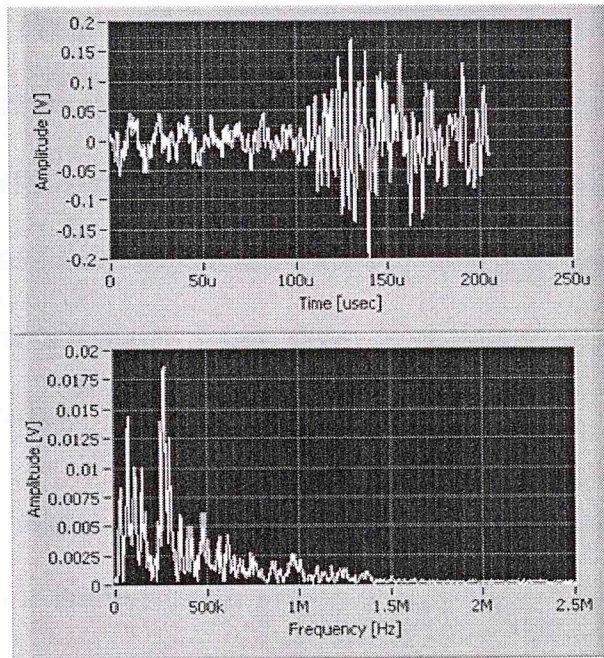
Figure 2.10 TOP: A "medium" signal from anAE event in a HF experiment in the time domain. BOTTOM: Frequency spectra using RMS magnitude.
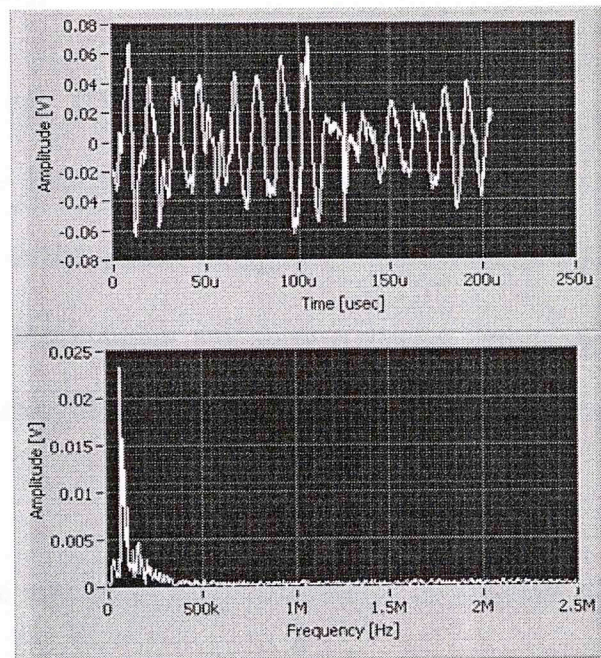
Figure 2.11 TOP: A "bad" signal from an AE event in a HF experiment in the time domain. BOTTOM: Frequency spectra using RMS magnitude.

2.2.5.1 Primary Frequency (Characteristic frequency)

Another characteristic of every signal is its primary frequency which contains important diagnostic information. I attempt to analyze the primary or principal frequency characteristic of events but these too vary with signal quality, as one would expect. For a "good" signal there is normally a dominant narrowly defined peak frequency.

On a "medium" signal the frequency spectrum is spread around different frequencies. Even if the maximum peak it is possible to be detected, this peak appears to be dispersed over different frequencies. For a "bad" signal, the frequency is also lower than a "good" signal. This is due to the reflection content on this signal.

Ohmaka and Mogi (1981) mentioned this difference of the frequency dependence of the quality of the signals in practice; one emission event differs from another in both the quality of signal observed and its frequency content. So according to this, every signal will have a unique frequency spectrum and the analysis of the primary frequency could characterize and differentiate one signal from another.

A summary of the maximum peak frequencies on real HF signals is shown on the following table:

Event Reference	SNR (dB)	Characteristic Frequency [kHz]
Good P6HF_E19S1	30.84	68.43
Good P10HF_E1S1	51.38	268.82
Good P15HF_E8S1	55.72	83.80
Medium	2.90	268.82
Bad	2.72	68.43

Table 2.9 Primary or characteristic frequencies from real HF events

It is difficult to find a pattern in the behavior of the characteristic frequency of the signals. As mentioned by Niwa et al. (1981), the frequency content of the AE waveforms depends on the cracking mechanism, the ray paths and attenuation in the sample. It is also possible to find a difference in the frequency behavior with different time periods in the sample experiment, this is mentioned by Sondergeld and Estey (1981) and Chitrala et al., 2010.

Analysis of primary frequencies of synthetic signals will be controlled by the principal frequency of the signal, which is I fixed at 350 kHz, and the amount of noise added to the signal. Table 2.10 shows this difference for different increments of signal to noise ratio (SNR).

Event Reference	SNR (dB)	Primary Frequency [kHz]
1 (Good)	(Inf) No Noise	350
2 (Good)	73.9	350
3 (Good)	63.2	350
4 (Good)	54.4	350
5 (Good)	42.8	350
6 (Good)	33.24	350
7(Medium)	23.44	350
8 (Medium)	11.9	345
9 (Noise)	8.56	345
10 (Noise)	6.45	345
11 (Noise)	3.44	335
12 (Noise)	1.34	350

Table 2.10 Primary frequencies for synthetic signals according to variation on SNR.

2.2.5.2   Time vs. Frequency analysis

Therefore, different ways to visualize and interpret the behavior of a waveform, the two most common are time and frequency domain. AE signals are commonly analyzed in either the frequency or time domain. Amplitude vs. time (time domain) shows how the amplitude of the wave varies positive or negative with time. Amplitude vs. frequency (frequency domain) shows the spectral

content or the frequency components of the signal and helps to highlight information that might be hidden in the time domain.

Another useful way to analyze the behavior of a signal over the time is using the short-time Fourier transform (STFT) spectrogram which provides the power (square of the magnitude) spectra over short time scales. The STFT is the simplest and computational fastest method to analyze a non-stationary signal (signal that frequency content varies over time). I compute the STFT using the following equation:

$$STFT[m,n] = \sum_{i=mdM-L/2}^{mdM+\frac{L}{2}-1} s[i]\gamma^*[i-mDM]\exp\left(-\frac{j2\Pi ni}{N}\right)$$

Where, $s[i]$ is the signal, γ is the window function, L is the window length, dM is the time step and N is the number of frequency bins. This function uses a sliding window to divide the signal into several blocks and applies the Fast Fourier transform (FFT) to each data block. Window tapers are useful in reducing spectral leakage, in our case a universal window called Hanning is used. The window function determines the resolution of the STFT. For better resolution a longer window length should be used but this causes a worse time resolution (LabVIEW Help, 2010).

Figure 2.12 (middle) shows an example of an STFT spectrogram on a "good" signal obtained from a real HF experiment using an automatic adjustment of the time steps that does not exceed 512 bins (recommended by NI.com) and a Hanning window with a length adjusted at four times the time steps. Hanning window helps to control the relationship between the time resolution and the frequency resolution of the time-frequency representation.

On a spectrogram is easily to identify different zones that correspond to the behavior of the frequency and amplitude of the signal over the time. Frequency is represented on the y-axis and the amplitude by the color density, while the x axis represents the time in μsec.
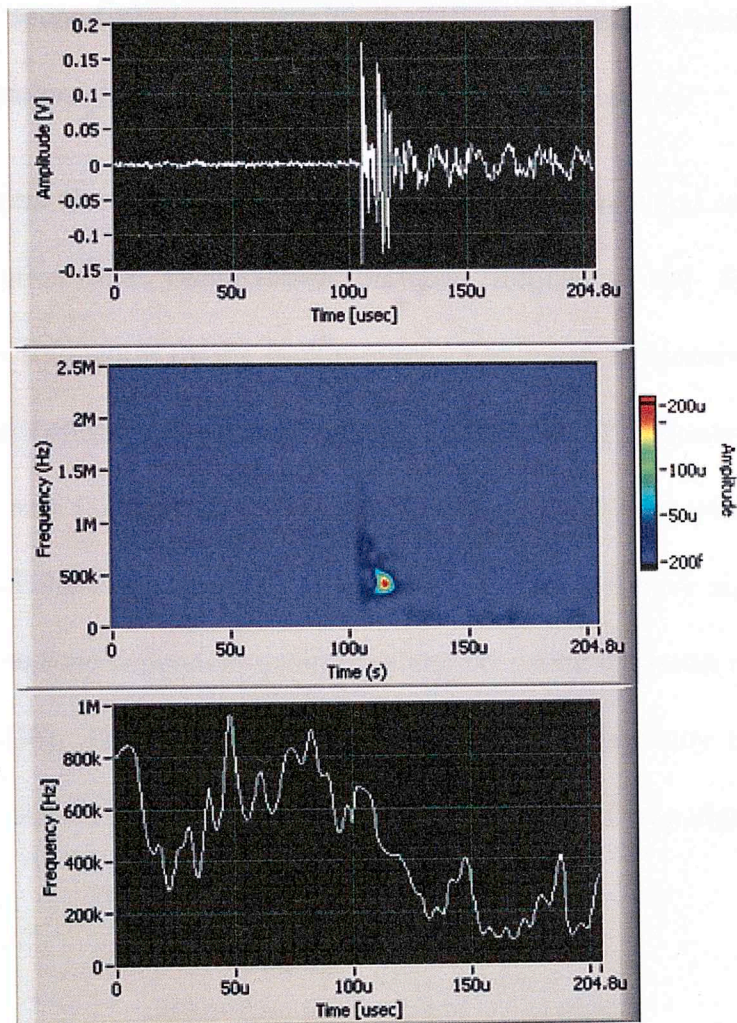
Figure 2.12 TOP: Time domain representation of a "good" signal. MIDDLE: STFT-spectrogram from a "good" signal obtained from a real HF event. BOTTOM: mean instantaneous frequency (MIF) obtained from the spectrogram showing the relation between frequency and time.

A "good" signal is expected to have a "quiet" zone which contains only low level background noise ($< 0.01mV$). The length of this zone depends on the selection of the pre-trigger point (for figure 2.12 the pre-trigger is selected at

100 μsec). A second zone will contains the HF signal which is recognized by a high concentration of colors and a different frequency behavior.

The spectrogram shows the relation between time, frequency and amplitude. For a single relation and visualization between frequency and time a mean instantaneous frequency (MIF) is computed. The mean frequency of a signal describes the central distribution of the spectrum. The spectrum of non-stationary signals (signals who changes frequency in function of time) is time dependent and therefore the mean frequency of non-stationary signals is time dependent. The time dependent mean frequency is called the mean instantaneous frequency (MIF). The MIF reveals how the central frequency of the signal changes over time. The following equation defines the MIF of a signal:

$$(\omega)_t = \frac{\int_{-\infty}^{\infty} \omega SP(t, \omega) d\omega}{\int_{-\infty}^{\infty} SP(t, \omega) d\omega}$$

Where $SP(t, \omega)$ represent the spectrogram of the signal and ω the frequencies involved on the signal. Figure 2.12 (bottom) shows an example of a MIF plot for a "good" signal. It is possible to clearly identify the different frequency zones and its change over time.

Before the arrival of the signal a zone with frequencies around 500 kHz up to 1 MHz represent the background noise. Following the arrival a reduction in the frequency is visualized indicating that the arrival has, in this case, a frequency

around 200 to 500 kHz. This behavior of the frequency of the signal is a characteristic that differentiate a "good" signal from a "bad" signal. Figures 2.13, 2.14 and 2.15 shows an example of another "good" signal obtained from a different experiment and a "medium" and "bad" signal obtained from real HF experiments respectively.
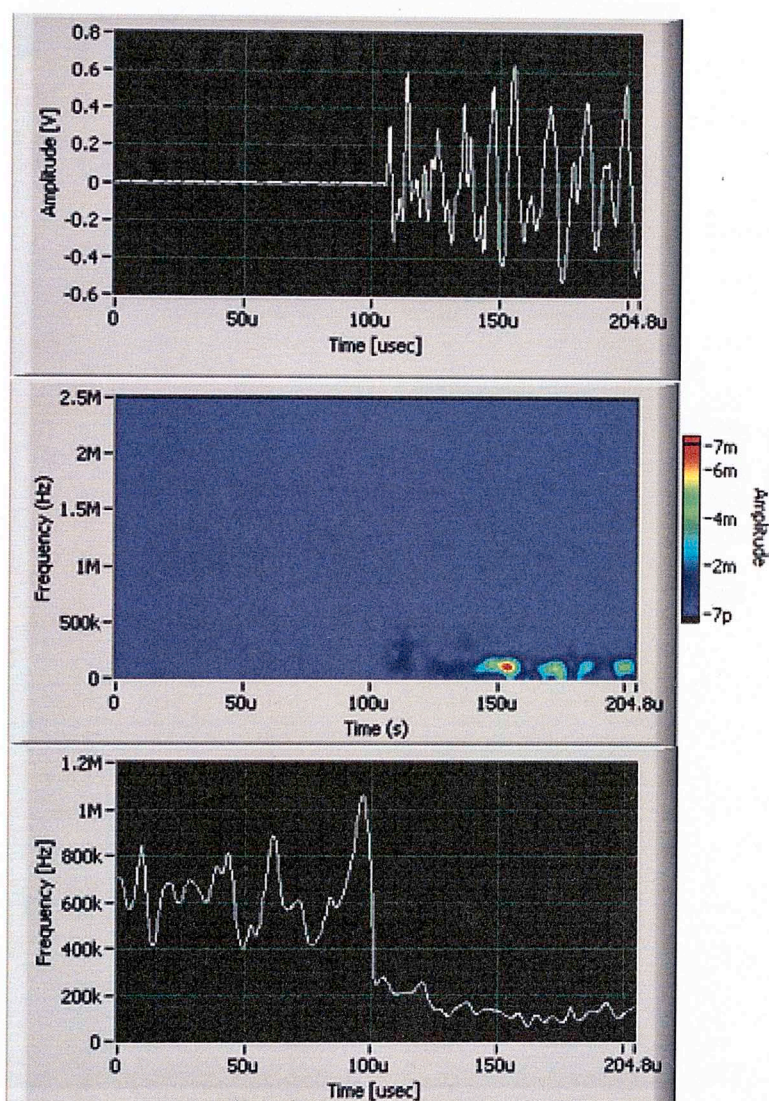
Figure 2.13 Example of a STFT spectrogram and MIF of a "good" signal obtained from a different experiment. The path mentioned for figure 2.12 is repeated.

Figure 2.13 shows an example of a "good" signal from a different experiment. The pattern that is mentioned for figure 2.12 is repeated and two different sections are distinguished; the "quiet" and the signal zone, before and after the pre-trigger point.
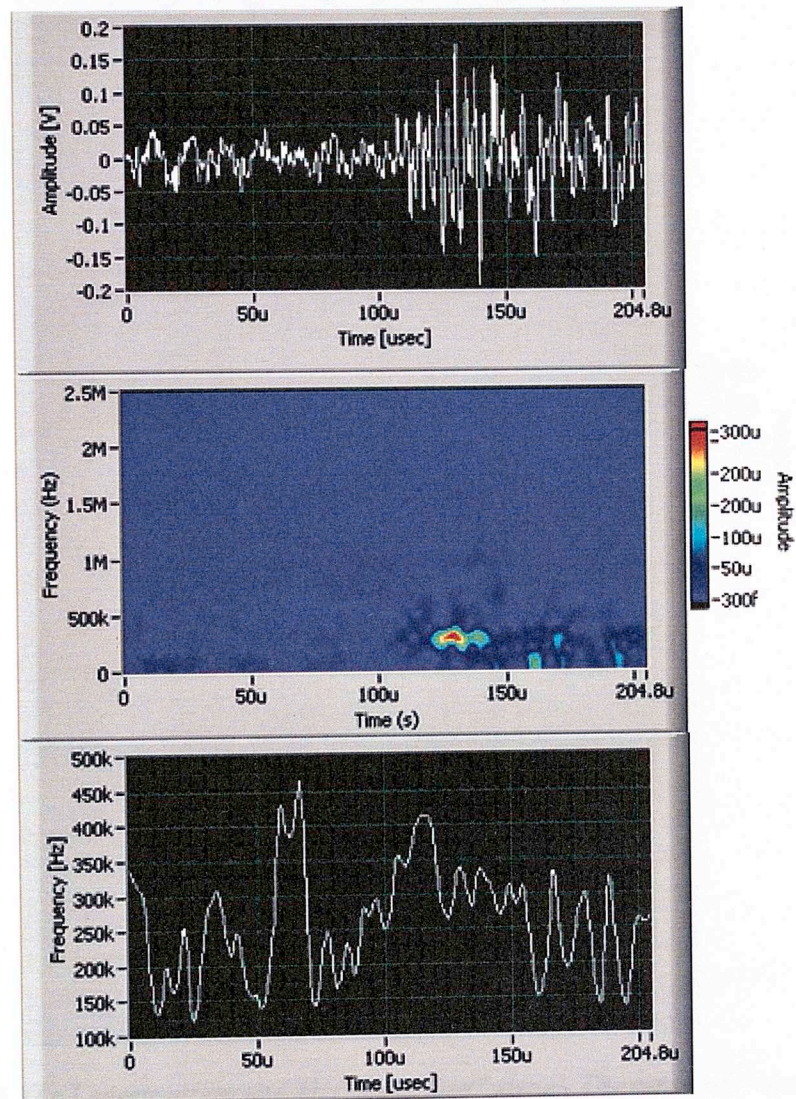
Figure 2.14 STFT Spectrogram and MIF for a "medium" signal. The complexity of the signal is evident on the MIF (bottom) plot.

Knowing the pre-trigger point it is possible to analyze the behavior of the frequency in the signal zone (where the arrival is expected) and perform a statistical analysis to obtain the mean, median and mode of the frequencies for this zone, in order to differentiate a "good" signal from a "bad" signal.
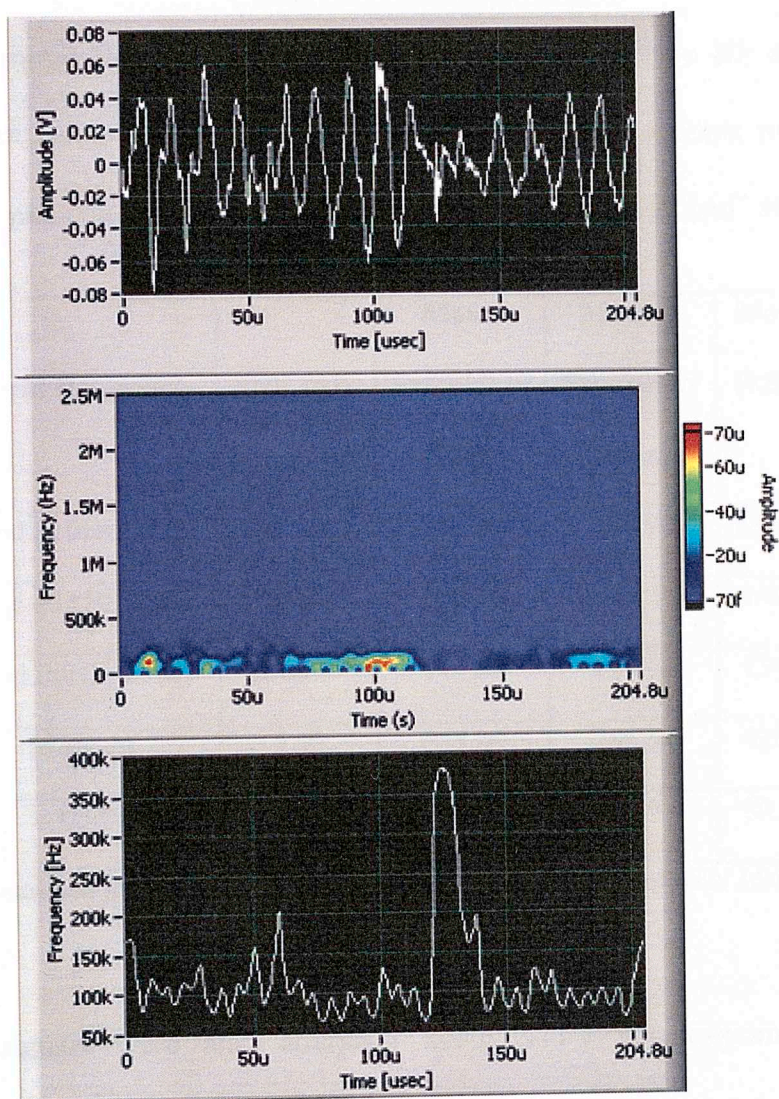
Figure 2.15 STFT spectrogram and MIF for a "bad" signal. The range of frequencies for this signal is evidently different from a good signal. The range of frequencies corresponds to the capture of a signal from a real event.

Analysis of MIF is a good indication that "bad" signals are not always obscured by background noise but by echoes and reflections of original signals around the sample.

Table 2.11 shows the statistical results obtained from MIF on HF signal from different experiments. It should be note that different experiments will produce different frequencies that differentiate a "good" signal from a "bad" signal.

Event Reference	SNR (dB)	Mean Frequency [kHz]	Median Frequency [kHz]	Mode [kHz]
Good P10HF_E1S1	51.38	343.84	289.92	289.33
Good P15HF_E8S1	55.72	205.93	167.9	146.4
Good P6HF_E19S7	30.84	146.08	128.08	134.9
Medium	2.90	278.09	285.51	412.2
Bad	2.72	126.3	96.9	80.15

Table 2.11 Statistical frequency analysis of the signal portion obtained from MIF of different types of signals from HF experiments.

Synthetic signals were also analyzed using STFT spectrogram and MIF statistical frequency analysis. Figure 2.16 shows a synthetic signal with a high SNR and is considered as a "good" signal. On MIF plot it is easy to identify the

different frequencies that represent the "quiet" zone (0 – 60 μsec) and the signal

zone (from 60 to around 120 μsec).
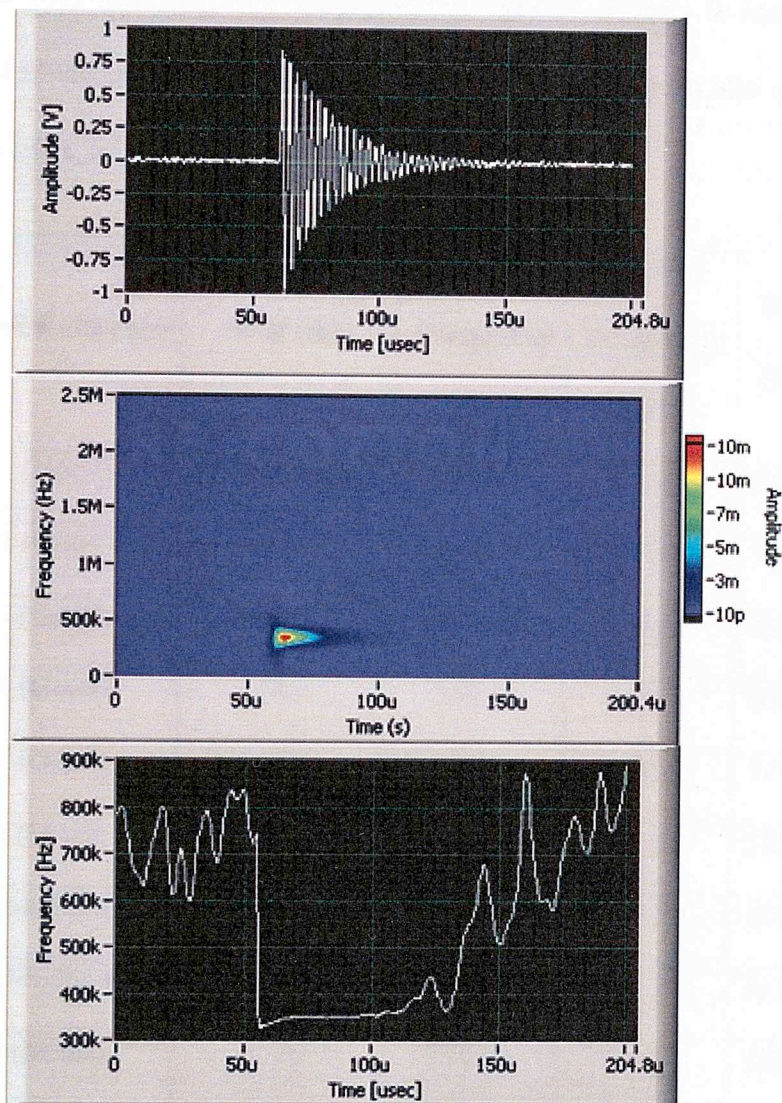
Figure 2.16 STFT spectrogram and MIF plot from a synthetic signal with good SNR.
(SNR=54.4 dB)

Table 2.12 shows the statistical analysis on MIF using synthetic signals. Analyzing the results from real and synthetic signals, it is possible to differentiate that normally the frequency of "good" signals is lower than the background noise, and higher than the frequency of reflection and echo signals, normally known as "bad" signals.

Event Reference	SNR (dB)	Mean Frequency [kHz]	Median Frequency [kHz]	Mode [kHz]
1 (Good)	(Inf) No Noise	348.07	349.01	349.88
2 (Good)	73.9	348.24	349.05	349.79
3 (Good)	63.2	348.72	349.59	350.47
4 (Good)	54.4	350.68	350.29	349.57
5 (Good)	42.8	352.35	351.78	349.90
6 (Good)	33.24	372.51	360.89	352.76
7(Medium)	23.44	407.90	373.58	351.98
8 (Medium)	11.9	493.78	475.08	363.37
9 (Noise)	8.56	576.83	550.26	359.62
10 (Noise)	6.45	587.14	503.62	459.23
11 (Noise)	3.44	723.11	654.61	616.34
12 (Noise)	1.34	742.73	740.73	704.48

Table 2.12 MIF statistical analysis for synthetic signals using different SNR levels.

2.3 Multidimensional Data Reduction

A combination of algorithms or attributes has the highest potential for consistent data classification (Tan, 2007). According to Tan (2007), statistical algorithms like threshold, histogram and zero-crossing count were found to yield the best results in classification. It is important to mention that the characteristics of the signals on the field are different with the ones obtained on lab-scale. This depends on the nature of the fracture and also on the equipment used to acquire the signals.

Different combinations of classification algorithms were tested to reduce the dimension of the data and simplify classification. Principal component analysis (PCA) is the multidimensional (also known as multivariate) data reduction technique used.

Multivariate analysis use statistical techniques that consider two or more random variables as a single entity (Jackson 1991). For this case, 6 different variables were used; three proposed by Tan, (2007): threshold, zero-crossings and histogram and three more: the signal to noise ratio (SNR) and 2 obtained from the frequency analysis (primary frequency and MIF mean frequency).

Tan concludes that using only three statistical algorithms (threshold, histogram and zero-crossing) yields the best results for classification. After reviewing our

results using different combination of algorithms, the best results were obtained using the threshold, mean of frequencies and the signal to noise ratio.

2.3.1 Principal Component Analysis

PCA is a way to identify patterns in data, and express the data in such a way to highlight their similarities and differences. The patterns that are expected in the data are generated by the different classification algorithms and using PCA to reduce the dimension of the data and finally make the decision of the difference between a "good" and a "bad" signal.

The dataset generated by an experiment contains many variables, which makes it difficult is to extract a single discriminator. Mathematical theory pertaining PCA is encountered on Labview Help, Smith (2004), Jackson (2004) and Tan (2007). PCA has been applied in many different fields including geophysics, for electrofacies characterization, computer science for image processing and pattern recognition, astrophysics and neuroscience.

2.4 Results

A dataset of microseismic signals taken from a HF experiment is used to analyze the performance of the classification algorithms and the final result of the classification. The standard results of the classification are determined by manual inspection. A single manual classification is made between "good", "medium" and "bad" events only for few events (for example lead break

calibration). On a large number of events, it is only possible to manually separate signals between "good" and "bad." It is necessary to remember that a "medium" signal is also considered as a "bad" signal, but a manual separation is also made because these signals represent the biggest challenge to manual classification.

A good event is known as a collection of at least 6 "good" signals acquired by the sensors surrounding the rock sample (normally 14 to 16 sensors are used in each experiment). If at least 6 signals are classified as "good" signals, the total event is also classified as a "good" event.

Tan (2007) suggested normalizing the algorithm's outputs for each signal of the dataset: "Normalization indicates that each algorithm measurement was divided by the largest value pertaining to that measurement over all examined microseismic files."

A sample of 30 events was extracted and organized in the following order: first 10 "good" events (160 signals), then 10 "medium" (160 signals) and finally 10 "bad" events (160 signals). A total of 480 signals (16 signals for every event) are analyzed and the results for every classification algorithm are presented on figures 2.17 to 2.22, where y-axis the represents the value obtained from the algorithm applied and x-axis corresponds to every signal for every sensor in every event.
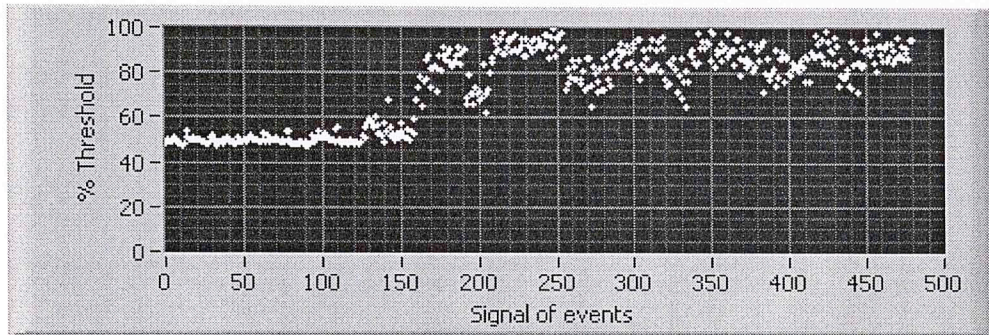
Figure 2.17 Normalized output from % of threshold value (value = 0.01V)
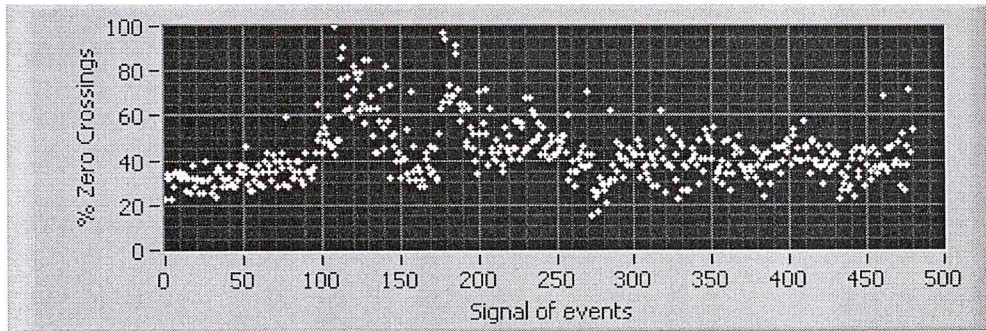
Figure 2.18 Normalized output from % zero-crossings (value = ± 0.03V)
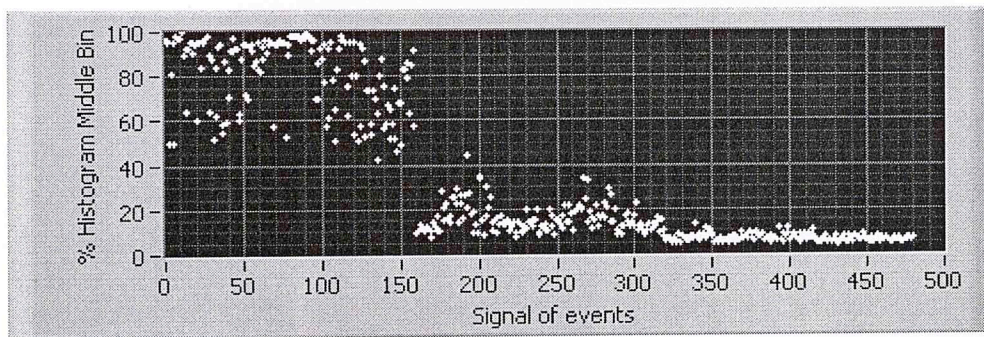
Figure 2.19 Normalized output from % of histogram middle bin (99 bins used)
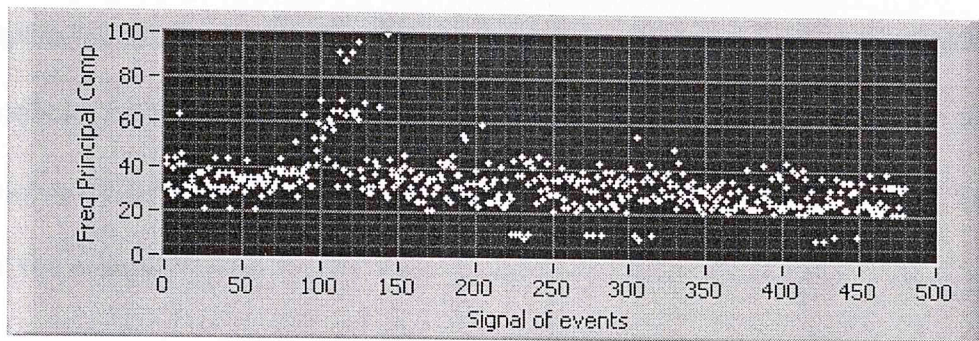
Figure 2.20 Normalized output from principal frequency component obtained from FFT plot
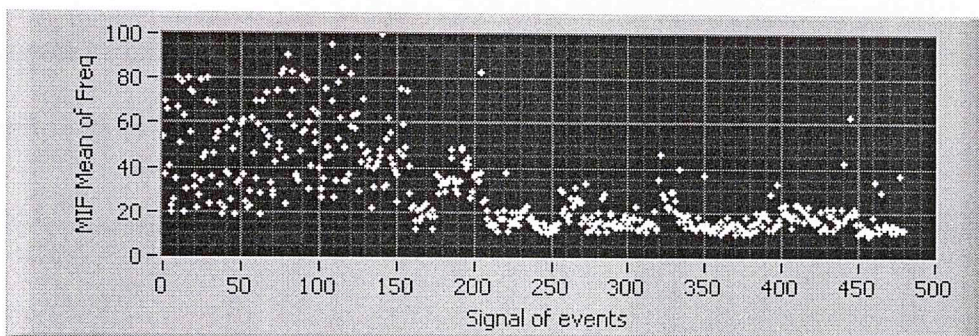
Figure 2.21 Normalized output from STFT spectrogram – MIF mean frequency
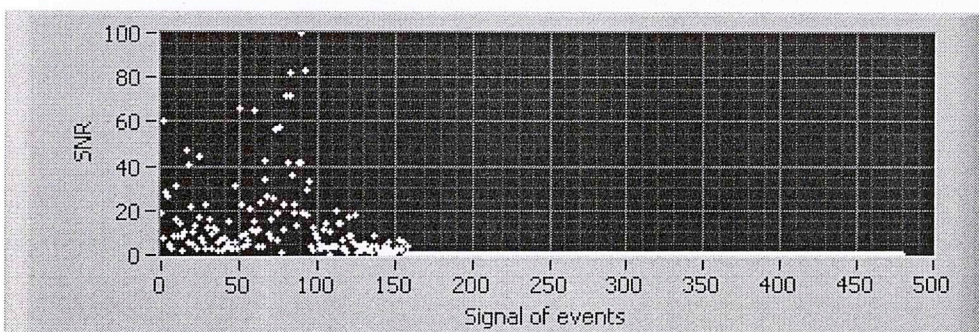
Figure 2.22 Normalized output from signal to noise ratio analysis (Noise window = 0 to 100 μsec) (Signal window = 100 to 200 μsec)

A clear pattern is observed in figures 2.17, 2.19, 2.21 and 2.22 corresponding to threshold, histogram, MIF mean of frequencies and SNR algorithms. The pattern observed provides a clear differentiation between the first 160 "good" signals and the others.

According to Tan (2007), the threshold, histogram and zero crossings, corresponding to statistics algorithms, are the ones that yield with the most accurate results in classification. Using PCA on the three algorithms outputs mentioned by Tan (2007), produces the results presented in figure 2.23.
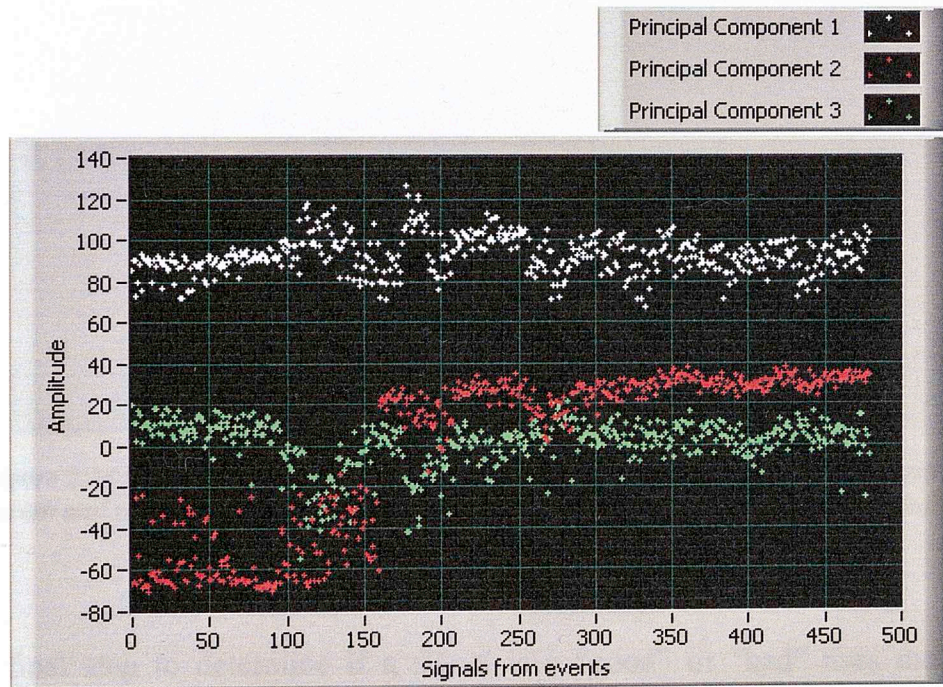
Figure 2.23 First, second and third principal components obtained from PCA applied on threshold, histogram and zero-crossing algorithms. Second and third components will be used to make the final decision and differentiation between "good" and "bad" signals.

It is clear that principal component 2, creates 2 zones that correspond to the "good signals" from 1 to 160 and "bad signals" from 161 to 480. Principal component number 3 defines the limit between a good and a bad signal. Finally, these two components will be used to differentiate a "good" from a "bad" signal. Figure 2.24 shows these 2 components plotted on the same graph.
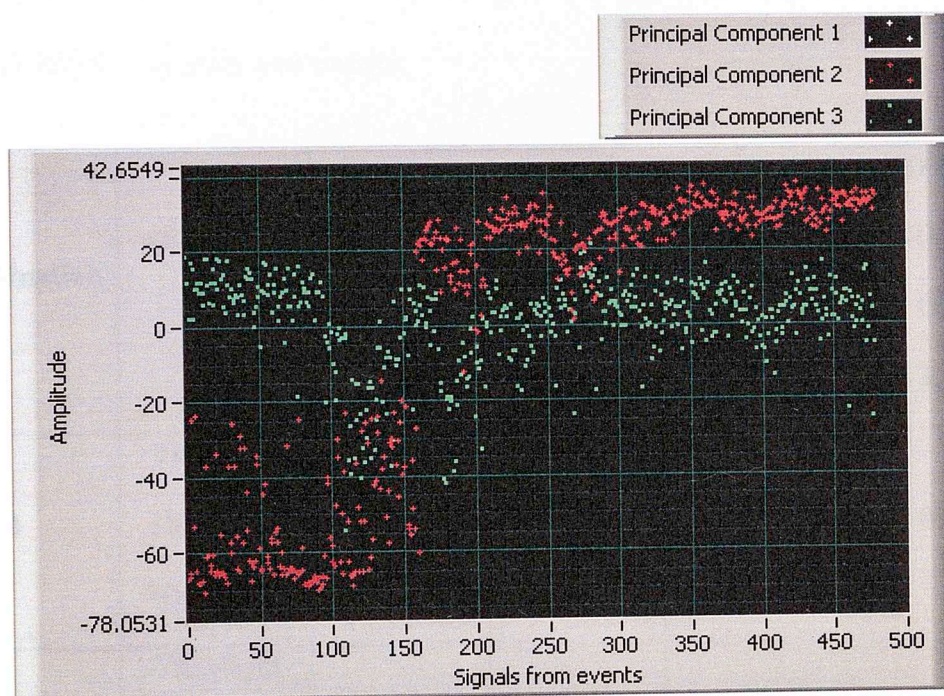
Figure 2.24 Second and third principal components using J.Tan algorithms (threshold, histogram and zero-crossings). Principal component number 3 limits a good signal from a bad signal.

The final step to determine if a signal is a "good" or "bad" uses the value principal component 2 (PC2), if this value is less than the one obtained from the principal component number 3 (PC3), the signal is considered as a "good" one, otherwise is considered as a "bad" signal.

$$if\ PC2\ \leq PC3\ \text{"GOOD"}\ signal$$

$$if\ PC2 > PC3\ \text{"BAD"}\ signal$$

Based on these expressions, PCA was applied on the different combinations on the different algorithms created. The best results are summarized in the following table. The error is counting in comparison against a manual classification of signals and events:

Combination	Algorithms Used	Error on Classifying Good Signals	Error on Classifying Bad Signals	TOTAL ERROR ON SIGNALS	TOTAL ERROR ON EVENTS
1	Threshold Histogram Middle Bin	6	0	98. 75%	0%
2	Threshold SNR Histogram	0	5	98.9%	0%
3	Threshold MIF SNR	26	0	94.5%	0%

Table 2.13 Total results from the best combinations of algorithms. All the combinations results with a 0% of error (comparison vs. manual classification) in event classification. Combination number 2 yields the best results in signal classification.

Figures 2.25 show the results from the second and third principal components using the combinations number 2 and 3.
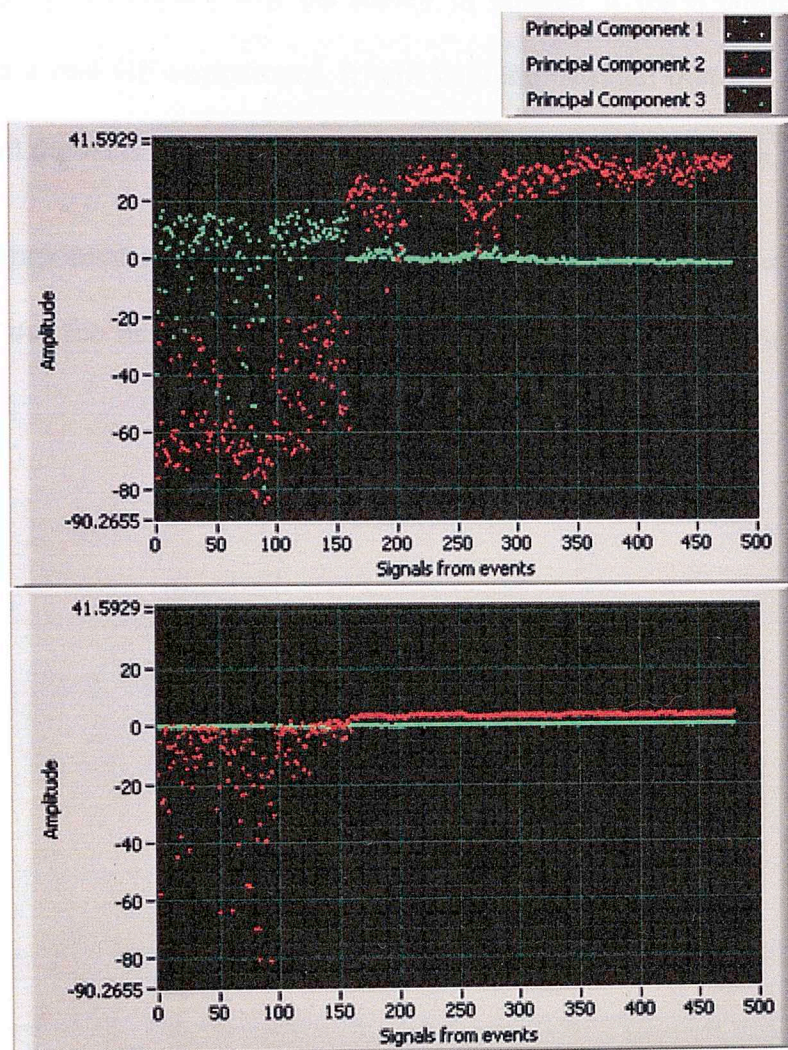
***Figure 2.25 TOP: Principal components 2 and 3 using the combination of algorithm 2.
BOTTOM: Principal components 2 and 3 using the combination of algorithm 3.***

Testing on a diverse number of signals like the one presented on figures 2.24 and 2.25 yields a perfect classification with no bad signals detection. This

perfection in not expected over all the datasets, but the demonstrated improvement after the application of principal component analysis is expected (Tan, 2007). Other results will be shown in chapter 5 for a complete set of signals from a real HF experiment. It will be demonstrated that the accuracy of classification depends on the diversity of the signal according to Tan, 2007.

Principal component analysis was first analyzed in all possible combination of algorithms, but the best results were obtained using only 3 combinations given in table 2.13.

3 ARRIVAL PICKING

One of the most important applications of AE data is determination of AE events location from observations at multiple stations (multiple sensors). Arrival time-picking is a critical step in the analysis of AE data and is required for the location of events. To obtain an accurate location of an AE event, two factors are important to know; velocity of the wave in the specimen and arrival time. This investigation will focus on arrival time picking from P-wave arrivals obtained on hydraulic fracture AE events.

After signal classification, only "good" signals should remain. These signals will be processed to extract arrival information. Due to the number of signals, an automatic method must be used to process the signal and extract this information. Numerous methods for arrival picking have been applied over the years

Energy, multi-window amplitude, S-transforms and autoregressive techniques in time and frequency domains are just a few samples of algorithms commonly used recently in lab scale. Their accuracies depend on numerous factors: amplitude, SNR and frequency of the signals, among others (Munro, 2005). Techniques as short term average vs. long term average (STA/LTA) and modified energy ratio (MER) emerge from the others for their recognized accuracy and fast convergence.

Using synthetic and real data, a comparison between manual and automatic pick have been carried out using STA/LTA and MER algorithms. Both STA/LTA and MER techniques work well when signals have good SNR ratios. However at lower SNR (SNR < 10), the STA/LTA algorithms fails (Wong et al., 2009), which makes MER the technique that was choose to finally pick all the arrival times of the signals classified previously.

Implementation of an automatic, fast and accurate method is highly desirable to process the amount of microseismic events. In order to get an accurate arrival, we implemented a process defined by R. Allen, which can be organized and structured depending on the type of signals used (Allen, 1982).
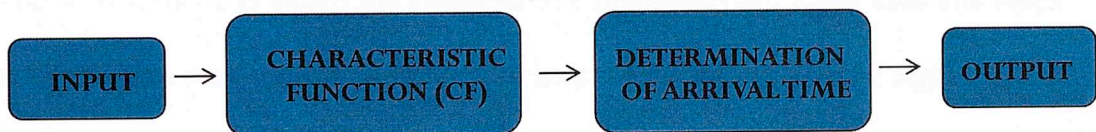
Figure 3.1 Block diagram of logical structure used for arrival pick of AE. (Adapted from R. Allen, 1982).

According to R. Allen, the logical structure of all pickers is strikingly similar and could be divided into a 4 block decision structure. The first block corresponds to the input section and includes different options to "smooth" the signal. Filtering and averaging can be applied in this point with the concern to preserve signal frequencies, in order to discard DC and high frequency components. R. Allen mentioned this block as the one that applies the input

filter and the analog to digital converter. It have been mentioned that these signals change its frequency with time, so the frequency we are interested in will be centralized around the arrival. Based on previous research performed on AE signals, using a window of 5 µsec about the first arrival, the principal frequency is between 150 and 547 kHz for an Indiana Limestone (Chitrala et al., 2010). So in order to preserve the important information, it is normally recommended to use a band-pass filter (50 kHz to 1.0 MHz). Several researches before have mentioned the concept: "the less filtering, the better the algorithm" (Douglas, 1997); for this reason an extensive analysis on filtering techniques will be avoided (Munro, 2005), but open for a future work

The second block is called the characteristic function (CF). Allen uses this block to generate a new time series characteristic of the digital filtered signal that is going to be examined for changes indicating the presence of an arrival. This is the part where the algorithm will be used (MER, STA/LTA, among others). The CF can be as simple as an absolute value of the filtered trace. The determination of which CF to use depends on the signal expected and the performance required of the picker.

For the third block another algorithm is used to finally make the determination of the arrival time. Allen declares that this block is used for the analysis on false results, but it is also mentioned the complexity of the implementation and also

the computer processing required for it. The decision could be made based on different thresholds values that are acquired after an analysis of hundreds of different signals (Allen, 1982). Threshold values will differ if the characteristics of the signals change and will make analyzing different samples using the same decision criterion algorithm impossible.

The decision of which criteria to implement in the third block depends on the CF algorithm used. The third block is called, determination of arrival time. In our case this block was used for a simple decision of arrival based on the CF. Algorithms implemented for this block will be discussed on every CF analyzed.

The final block is called output post-processing. Allen uses this block for verification analysis of a tentative pick. A good algorithm minimizes the number of steps required for a final decision. In our case this block is simply used as a final presentation of the information, a table showing the result of only the events and sensors previously classified as "good". The output is saved as *.LVM file whose format is compatible with Microsoft Office (Word and Excel).

The final evaluation of picker performance is carried out using a direct comparison of automatic and manual picks on real and synthetic data. The number of comparison with real data will be restricted to a few signals due to

the known error introduced by a manual hand pick on a big large number of signals.

3.1 Arrival Picking Algorithms

The arrival picking algorithm is implemented in the block 2 of the logical structure, which is called the characteristic function (CF). The final performance of the complete picker depends heavily on the choice of algorithm. An arrival is identified by a sudden change in amplitude or frequency, or both, in the time sequence series. The CF must respond to these changes as rapidly as possible, and, ideally should enhance the change (Allen 1982).

The majority of automatic pickers are designed for online detection and timing of primary wave (P-wave) arrivals (Earle et al., 1994) (reviewed by Allen, 1982). Our goal is to implement a method that works after the hardware acquisition, detection and software classification of the signals. This is known as an offline method for detection where all the processing and analysis is done after the signals are acquired.

3.1.1 Previous work

A brief review of previous work will be discussed in this part of the chapter. Some algorithms will be tested and described in detail, others will simply be mentioned.

Absolute value $CF(i) = |X(i)|$ is the most widely method used for detection of AE events and arrival picking even for signals with extremely low SNR. It is easy to compute but not viable to be used as a stand-alone arrival picker (Allen, 1982).

Energy analysis differentiates and enhances the SNR applying the following algorithm, $CF(i) = |X(i)|^2$, in other words, energy algorithm enhance the amplitude differences of the signal, but not the frequency changes.

Figure 3.2 shows an example of absolute value and energy algorithms applied for a real "good" AE signal. The arrival is visible; the absolute value function preserves the clean onset of the energy arrival which facilitates threshold picking. The direct method of absolute value is clearly not useful when a signal is dominated by noise; even at high amplitude levels the arrival could be lost.

Figure 3.3 shows how these methods are used for a "medium" class signal. In this figure the enhancement in amplitude of the energy method is better than the absolute value method (Bottom plot in figure 3.3).
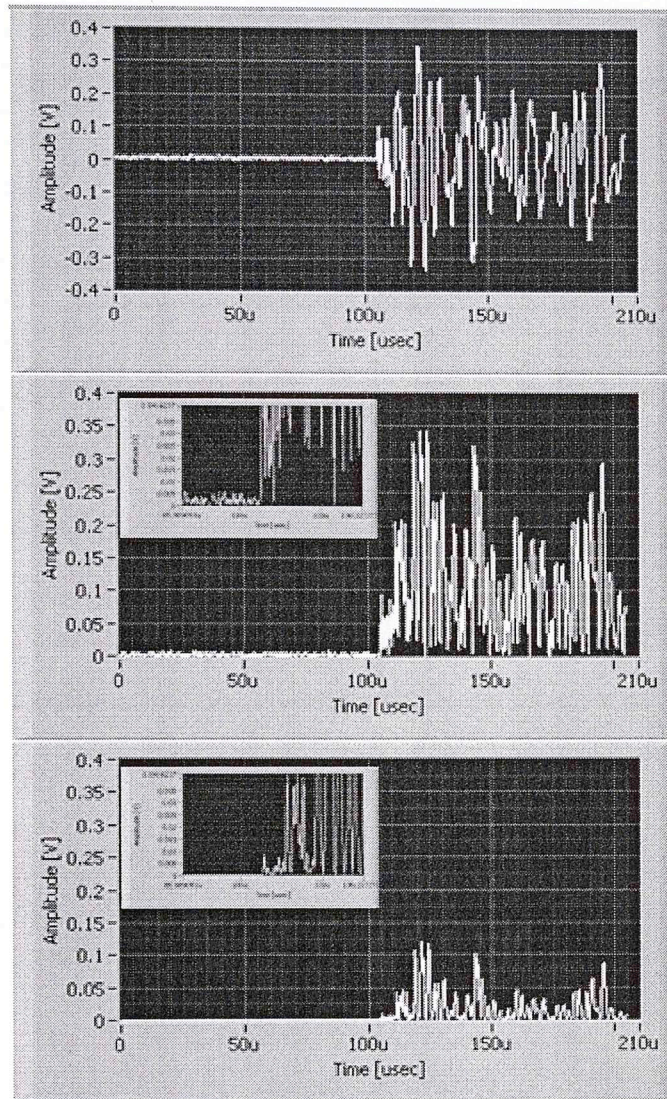
Figure 3.2 TOP: real AE "good" signal obtained from a HF experiment. MIDDLE: Absolute value algorithm applied to the time series. BOTTOM: Energy algorithm applied to the same time series.

Using the simple energy algorithm, it is possible to see a clear differentiation between signal and noise. This enhancement will be discussed later with STA/LTA and MER algorithms that also use this advantage.

74
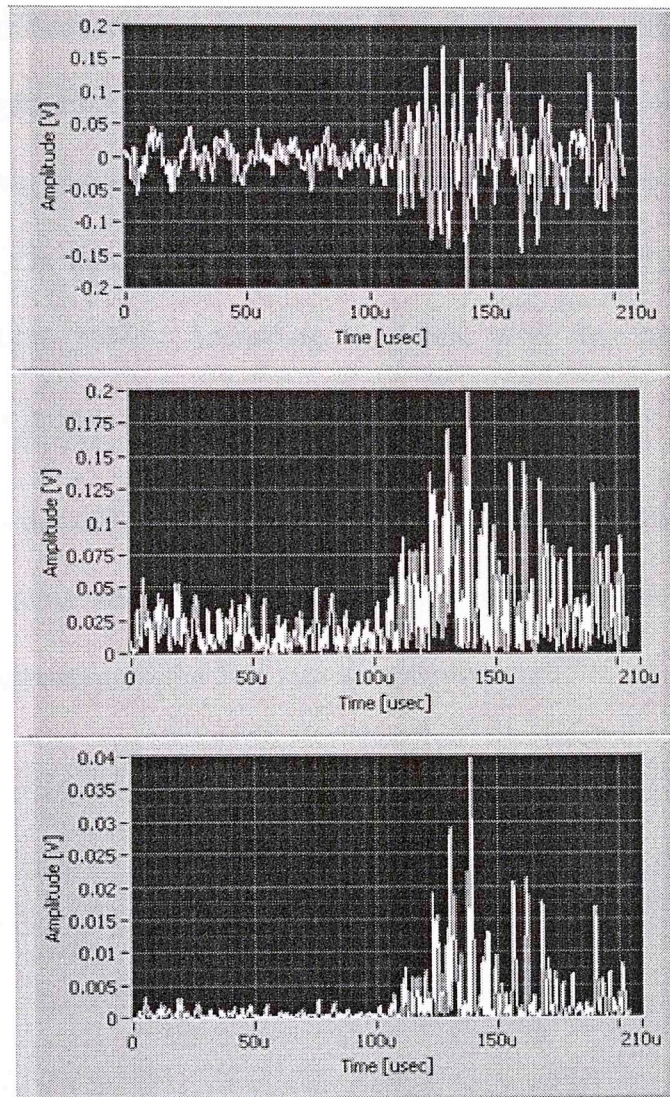
Figure 3.3 TOP: Real AE of a "medium" signal obtained from an HF experiment. MIDDLE: Absolute value algorithm applied to the original signal. BOTTOM: Energy algorithm applied to the signal. The energy algorithm provides a superior discriminator.

Stewart (1977) uses a function that analyzes the differences in the incoming signal. This algorithm really tracks polarity changes, $DX_k = X_k - X_{k-1}$ where k represents the current time. The sign of the difference is compared with the

previous first difference and so on and the sign of this difference is taken (positive or negative). The signs are compared for a length of time, if the signs persisted for less than eight consecutive times, then the value of the modified signal (MDX_k) is taken to be its current value increased by DX_k, (Stewart, 1977) and (Allen, 1982). According to Allen, with this technique slightly emergent onsets have a chance to be detected.

Figure 3.4 plots the difference algorithm by Stewart. A clear enhancement of the signal is not visible and right arrival time is questionable, but the oscillatory nature and the direction of the first motion are preserved.
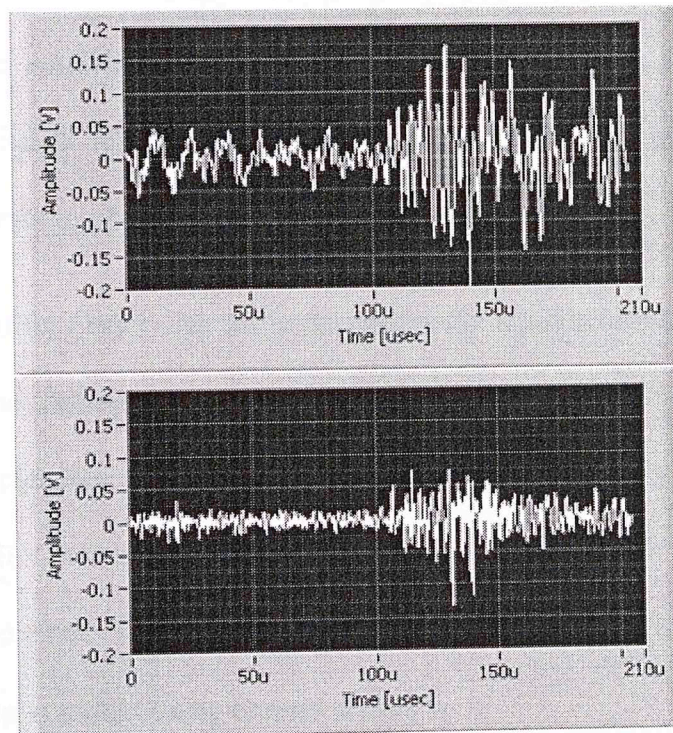
Figure 3.4 Stewart (1977) one sample algorithm applied on a medium class signal (bottom). A clear improvement in the ability to pick the arrival is not evident; however, some original characteristics of the signal are preserved.

After Stewart's algorithm is applied, an event declaration criterion needs to be used. The criteria use the absolute value of MDX_k. For a final declaration time, a set of criteria and thresholds must then be met. The prime advantage of Stewart's function is that it is very fast to compute. It only involves addition, subtraction and comparison operations (Allen, 1982).

Allen (1978) developed a detector based on an envelope that is equal to the square of the data plus the weighted square of the first derivative.

$$E(t) = f(t)^2 + C_2 + f'(t)^2$$

C_2 is a weighting constant, whose purpose is to vary the relative weight assigned to the amplitude and first difference depending on sample rate and noise characteristics. The processed data then is subjected to a set of logical and mathematical tests to obtain a final arrival pick (Withers, 1998).

Baer and Kradolfer (1987) started to transform the signals based on the work of R. Allen. Some deficiencies were detected for the specific type of signals analyzed. Deficiencies like the strong sensitivity to frequency increase rather than amplitude which is characteristic on tele-seismic and regional seismological events. To overcome these deficiencies, a new CF similar to Allen's was implemented using energy analysis.

$$e(t)^2 = y(t)^2 + y(t)'^2 * \frac{\sum y(t)^2}{\sum y(t)'^2}$$

This function is called the envelope function (EF). Where y(t) is signal and y(t)' is the derivative of the signal. The sum is taken from the beginning of the signal.

The final characteristic function (CF) is produced by a statistical test using the running average and variance of the function $= (e(t)^2)^2$, called the squared function (SF) in the following algorithm:

$$CF(t) = \frac{(SF(t) - \overline{SF(t)})}{S(t)}$$

Where, $\overline{SF(t)}$ is the average of the SF, and S is its variance taken from the beginning of the series to the present point (Baer and Kradolfer, 1987).

After the CF is obtained criteria are established to extract the arrival time. These criteria depend on the CF being analyzed. A change in the principal characteristics of the signals (amplitude, frequency) will require a change in the criteria applied.

Figure 3.5 shows an example of a CF obtained from a real HF "medium" signal using the Baer and Kradolfer algorithm. The signal contains a clear change in frequency which is visible in the previous figure 2.14. Here, the enhancement in

the CF is in the difference of amplitude where the peaks are obtained in the time

position of the largest amplitude peaks of the original signal.
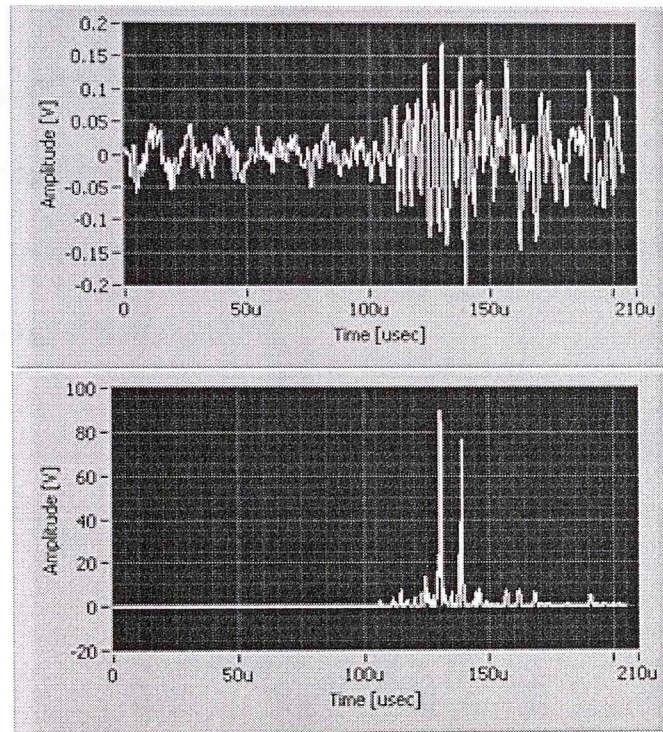
Figure 3.5 BOTTOM: Characteristic function obtained using Baer and Kradolfer algorithm. This algorithm is sensitive to changes in amplitude. The large peaks correspond to the largest amplitudes of the original signal.

To summarize, previous work can be divided or categorized into time and

frequency domain, particle motion and pattern matching, among others

(Withers, 1998). None of all the picker algorithms yield good results under all

situations. Most of the pickers widely in use today are based on time domain

methods.

Time domain methods are applicable to a larger range of signal types. The frequency of the signals does not behavior similar to the amplitude, and a manual picking is easier when the amplitude change rather than the frequencies.

3.1.2 STA/LTA

Short term average and long term average (STA/LTA) is classified as a time domain energy technique. The STA/LTA algorithm evaluates the ratio of short to long term energy density $(y(t)^2)$.

$$STA = \frac{1}{ns} \sum_{j=i-ns}^{i} y_j^2$$

$$LTA = \frac{1}{nl} \sum_{j=i-nl}^{i} y_j^2$$

$$if \; j <= 0, set \; y_j = (y_1 + y_2)/2$$

Here, ns is the number of data points (number of samples) taken for the window of the STA and nl the number of data points in the LTA. The long term average (LTA) characterizes the slow trend of signal energy, while the short term average (STA) is more responsive to sudden increases in energy (Oye and Roth, 2003).

The ratio STA/LTA is used as a measure of the signal to noise ratio (SNR). When the ratio exceeds a pre-defined constant threshold, a detection time is assigned to that specific signal (Munro, 2005). Trying to avoid certain criteria necessary to make a final time decision of an arrival, Wong et al., (2009) form the ratio from the derivative obtained from:

$$d_i = r_{i+1} - r_i, i = 1, 2, \dots (N - 1)$$

Where s represents the ratio STA/LTA. The maximum value of the numerical derivative of the STA/LTA ratio is close to the first-break time of the first arrival (Wong et al., 2009).
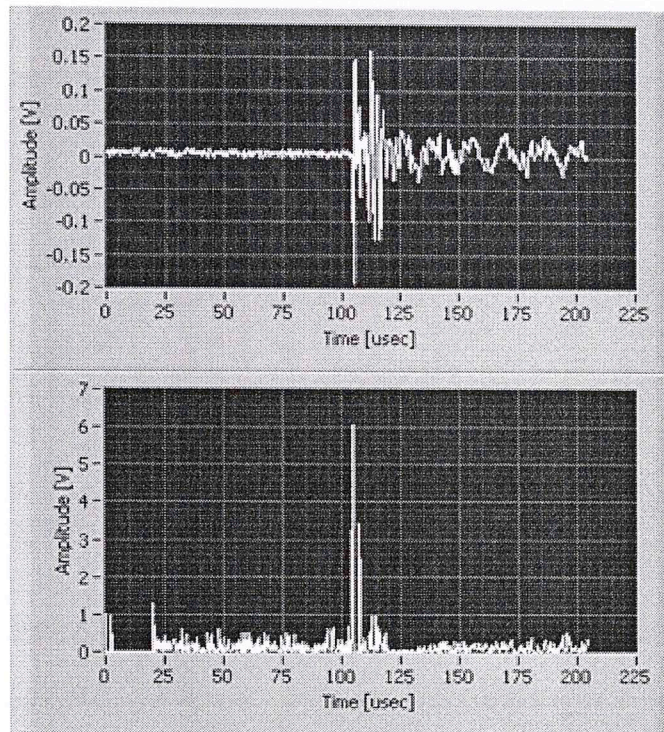
Figure 3.6 STA/LTA ratio applied to real HF "good" signal with a high SNR. The peak of the derivative of STA/LTA ratio corresponds to the arrival time of signal. Window size 10/100 for the sort and long term averages.

This method works well on signal with a high SNR (see figure 3.6), giving accurate results compared with manual result. If a signal has a low SNR (figure 3.7), the algorithm begins to fail. Other criteria will be required to define the correct arrival of a signal.
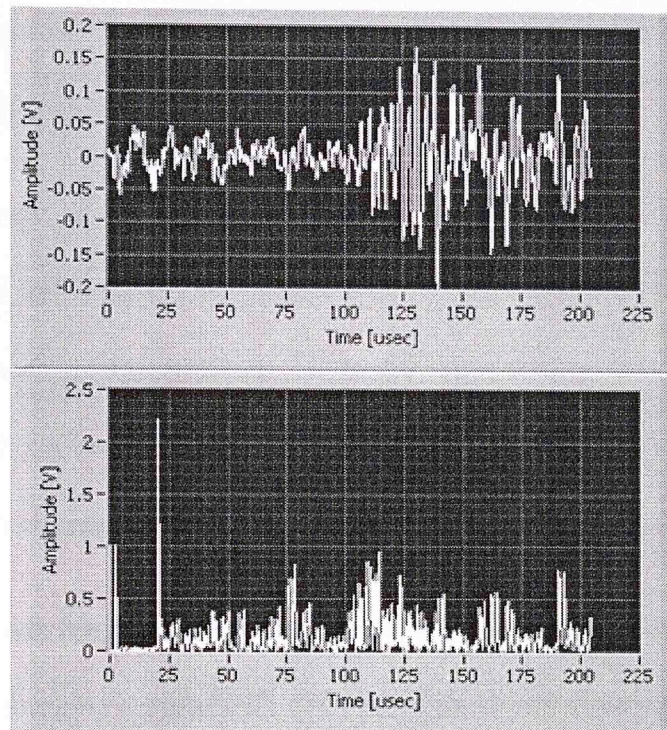
Figure 3.7 STA/LTA ratio applied to real HF "medium" signal with a low SNR. The peak of the derivate on STA/LTA ratio do not corresponds to the arrival time of signal due to the noise presented on the signal. Window size 10/100 for the sort and long term averages.

The pre-trigger points (number of points before the threshold crossing) could generate false arrivals due to strong noise levels. Some of these pre-trigger data points could be erased to prevent this error and obtain more accurate arrival results but in more of the cases this type of signals will be classified as a "bad" signal. This presumes that only "good" signals arrive after the pre-trigger. This is generally true if the arrival on any of the 16 sensors triggers the acquisition

The lengths of the windows used for STA/LTA depend on the type of signal. The STA is usually longer than a few periods of the typical expected signal, and

the LTA is longer than a few periods of typically irregular seismic noise fluctuations (Munro, 2005).

The typical range of frequencies expected for good signals is around 100 – 500 kHz. The final value for STA window was chosen between 2 to 10 μsec by trial and error, which corresponds to 10 to 50 data points and 500 to 100 kHz respectively. LTA (nl) window size has to be longer than STA window (ns) different window sizes were analyzed; a 100 μsec window gave the best results.

Figure 3.8 shows the manual pick of an arrival of a signal classified as a "good" signal which contains a high SNR (28 dB) without any filtering. This signal will be used to analyze and compare the response of the CF to different STA/LTA ratios. The results are presented on figures 3.9 to 3.14. The maximum peak of the CF in the STA/LTA ratio is closer to the arrival signal but a correction needs to be applied.
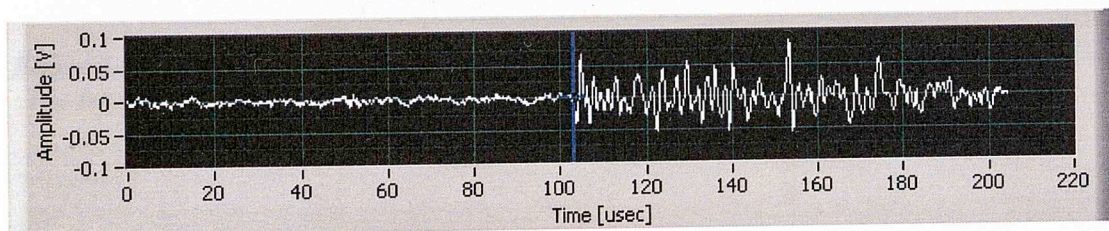
Figure 3.8 Manual pick of a "good" signal used to demonstrate the CF obtained from different STA/LTA ratios. The arrival is picked manually at 103.21 μsec. Automatic pick is sensitive to the window lengths used in the STA/LTA algorithm (see Figures 3.9 -3.14 below).
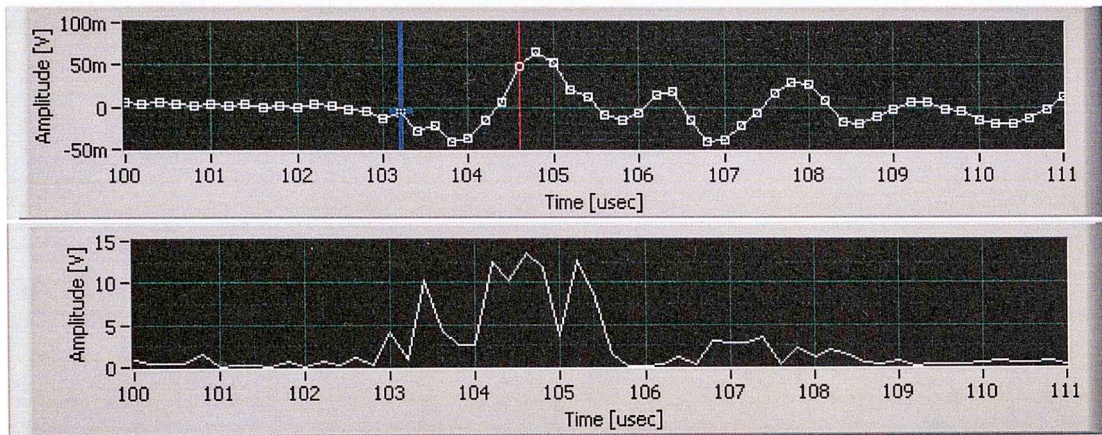
Figure 3.9 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 2/100. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak.
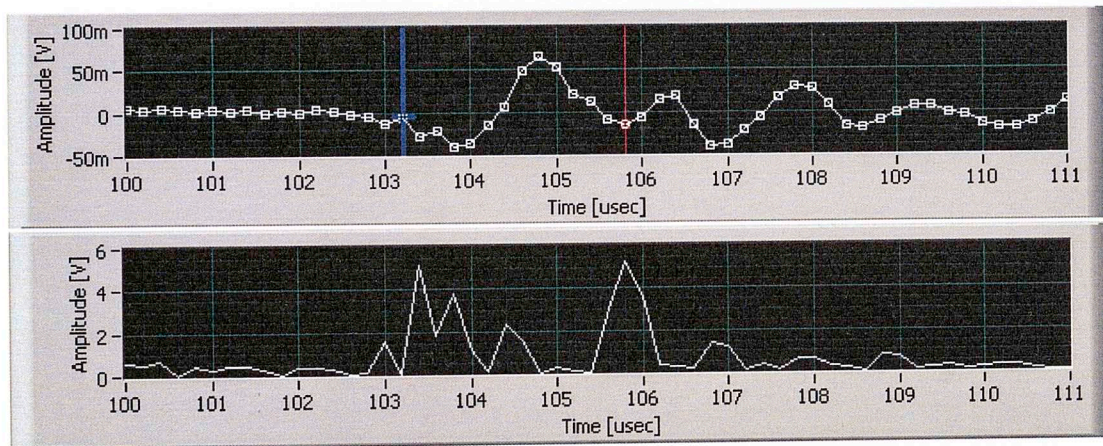
Figure 3.10 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 5/100. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak.
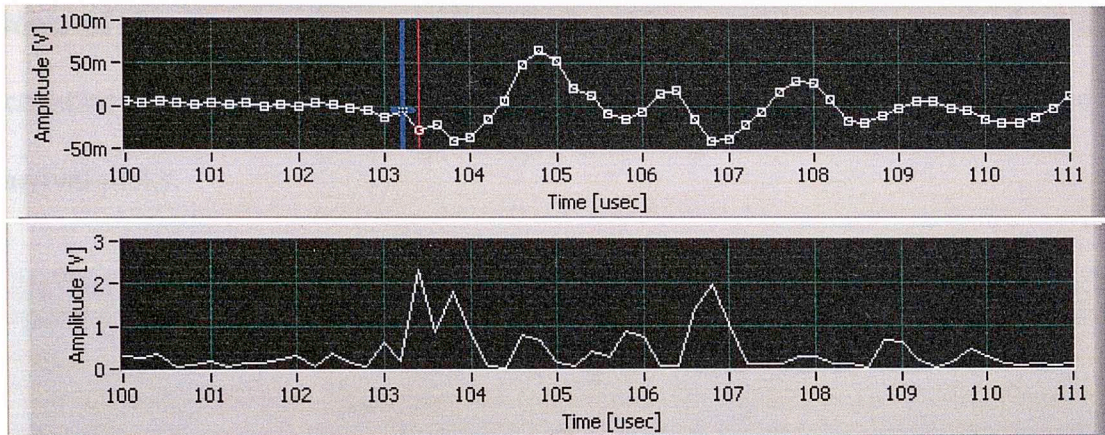
Figure 3.11 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 10/100. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak.

Figure 3.12 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 10/50. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak

The results obtained for the different values of STA/LTA window ratios, and the manual comparison with the original signal (figure 3.8) are summarized in table 3.1. A result in Table 3.1 indicates that the best results were obtained using an STA/LTA of 10/100 with an error of 3 data points. This error is constant for

most of the signals with a good SNR; for this case this error was used as a constant correction to minimize the difference between automated and manual arrival picks.
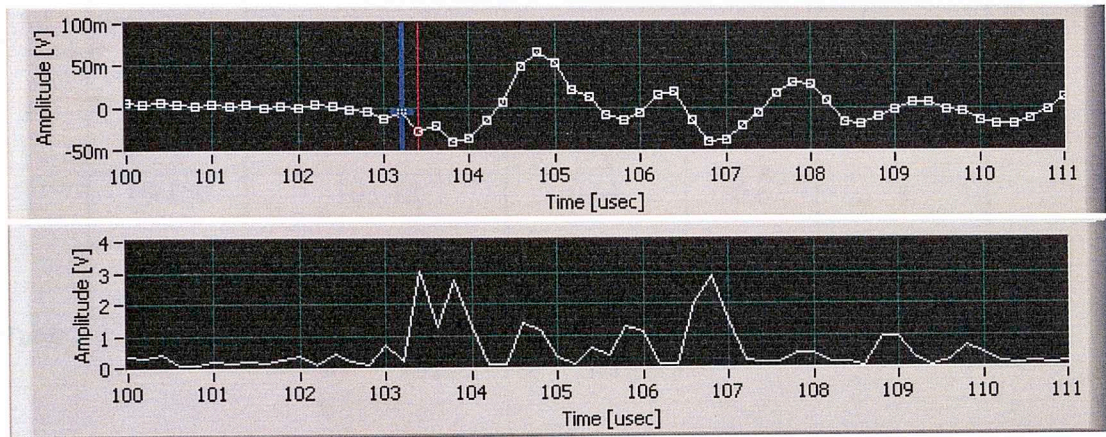
Figure 3.13 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 10/150. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak
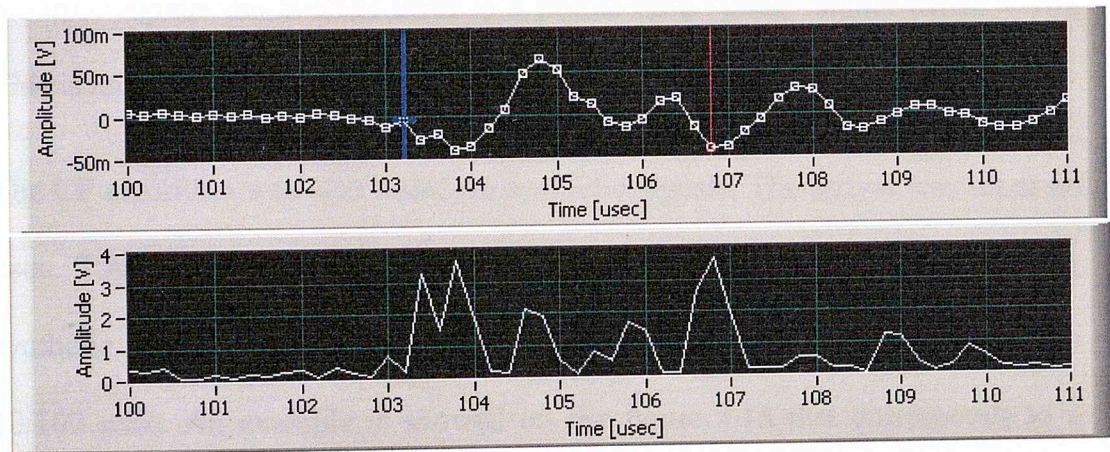
Figure 3.14 Manual (blue line) vs. automatic (red line) pick using an STA/LTA ratio of 10/200. The bottom plot is the characteristic function obtained from the STA/LTA ratio showing the maximum peak

STA/LTA	Automatic Arrival Pick Time [μsec]	Absolute Error [μsec]	Relative Error %	Data points difference
2/100	104.63	1.42	3.45	7
5/100	105.82	2.61	2.53	13
10/100	103.42	0.61	0.6	3
10/50	103.42	0.61	0.6	3
10/150	103.42	0.61	0.6	3
10/200	106.77	3.56	3.45	17

Table 3.1 Error analysis using different STA/LTA window ratios. The manual arrival pick is measured at 102.81 μsec

The results show that the minor error is obtained using an STA window of 10 μsec and LTA window of 50 to 150 μsec. After several trial and error tests on multiple signals the average error is 3 points on a "good" signal (SNR > 10). This is an acceptable average for arrival determination using this type of CF.

The CF algorithm was also tested on synthetic signals. The arrival was set at 60 μsec. Table 3.2 resumes the error produced by the automatic picking using synthetic signals with a fixed arrival in 60 μsec and an STA/LTA ratio set on 10/100 μsec. An example is showed on the figure 3.15 that corresponds to a signal with a SNR equal to 11.9 dB.

Signal - SNR	Automatic Arrival Pick Time [μsec]	Absolute Error [μsec]	Relative Error %	Data points difference
(Inf) No Noise	60.20	0.2	0.33	1
73.9	60.20	0.2	0.33	1
63.2	60.20	0.2	0.33	1
54.4	60.20	0.2	0.33	1
42.8	60.21	0.21	0.35	1
33.24	60.25	0.25	0.42	1
23.44	60.30	0.3	0.5	1
11.9	60.25	0.25	0.42	1
8.56	60.47	0.47	0.78	2
6.45	53.80	6.2	10.33	31
3.44	96.61	36.61	61.02	183.05
1.34	24.22	35.78	59.63	178.9

Table 3.2 Analysis of error using STA/LTA window characteristic function (10/100) with SNR variation on a synthetic signal with an arrival set at 60 μsec. Note the approach works well down to about a 10% SNR.
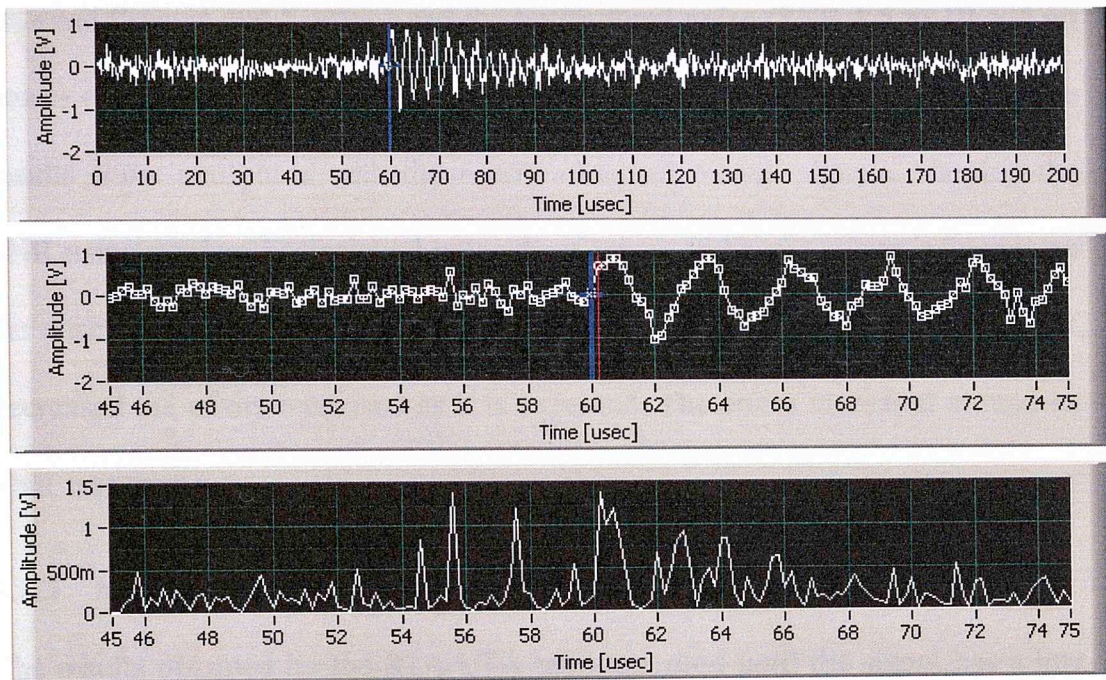
Figure 3.15 TOP: original synthetic signal with arrival at 60 μsec (blue line). The following plot corresponds to a zoomed zone around the arrival with an automatic arrival detected at 60.2μsec (red line). The last plot (bottom) corresponds to the CF of the STA/LTA ratio 10/100.

All the results shown to the point use neither any filtering technique to smooth the signal nor onset time correction. However, onset time correction is normally used on all the techniques and is based on the manual comparison error results. The difference in the number of data point between manual and automatic results are averaged and used for onset time correction. Complete results will be discussed on real HF signals after the classification is made in chapter 5.

Munro (2005) analyzed the behavior of STA/LTA algorithm without onset time correction and used a different technique for final picking. He analyzed the dependency of the frequency (20 Hz to 400 Hz), the amplitude (difference

between P-wave and S-wave) and the background noise (SNR 1.5 to 10 and no noise) on picking. Even if the characteristics of the signals are different, some results show similitude with the results obtained on our signals. Varying the SNR using random background noise does appear to have a minor influence on the onset time errors, with minor fluctuations of 1 or 2 data points. As the SNR decreased the errors increased as it is expected. The errors increased around an SNR less than 6.

3.1.3   Modified Energy Ratio

The results obtained by the STA/LTA ratio are good until the signal has a low SNR value. Wong et al., (2009) developed a time-picking scheme for individual traces based on a modified energy ratio (MER) attribute, which produced good results on signals with significant noise (Wong et Al., 2009).

The technique is based on the same energy STA/LTA ratio theory with variations on the window lengths. MER uses only one variable for the window size minimizing the probability of error by using two windows employed in the STA/LTA ratio technique.

Wong et al., (2009) described the method based on the following algorithm,

$$er = \frac{\sum_{j=i-ne}^{i} x_j^2}{\sum_{j=1}^{i+ne} x_j^2}$$

$$if, j \leq 0, set\ x_j = \frac{x_1 + x_2}{2}$$

$$if, j > N, set\ x_j = \frac{x_{N-1} + x_N}{2}$$

Where, x_j represents the time series of the seismogram with the time index $i = 1,2,\dots N$. The number of points in an energy time window is ne. And, finally the ratio of energies in windows preceding and following the index i is given as er (Wong et al., 2009).

$$er3_i = \left(abs(x_j) * er\right)^3$$

The peak of the modified energy ratio (MER) is taken over $er3_i$ algorithm. This peak is close to the first break on noise-free seismograms as shown in figures 3.16 and 3.17 for a high SNR and low SNR signals.

From the result obtained form a good signal (figure 3.16) it is clear that the maximum peak obtained from the MER algorithm corresponds closely to the arrival of the signal. In comparison with the result obtained on STA/LTA ratio, MER generates a clear peak signal with a clear arrival. This is a general characteristic produced by this algorithm when it is applied on a "good" signal with a high SNR.
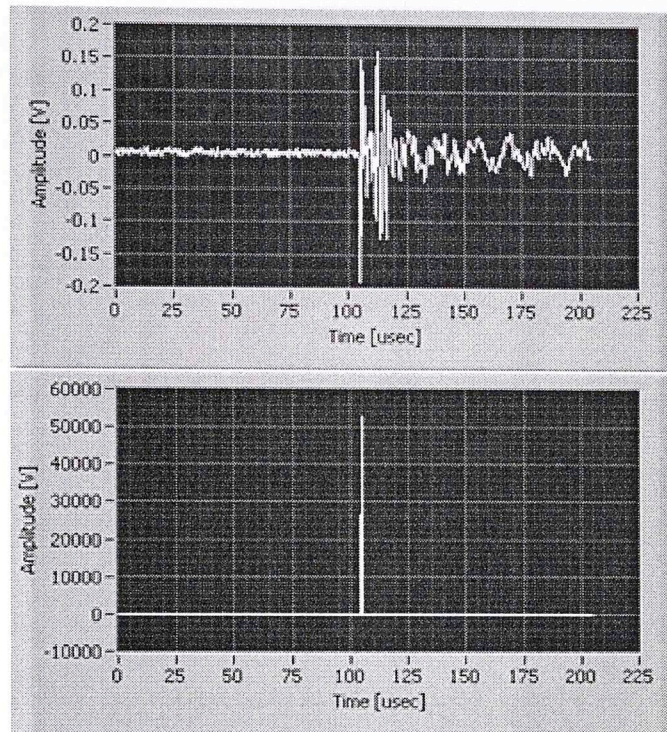
Figure 3.16 Modified energy ratio (MER) applied to a high SNR "good" signal obtained from a real HF experiment. Window size 100 μsec.

The major improvement of MER algorithm is the response on signals with a low SNR. Figure 3.17 shows an example of a "medium" class signal with an SNR = 2.9 dB; the maximum peak of the MER algorithm is closer to the possible arrival time of the signal. This wave corresponds to a signal that is difficult to manually pick due to the noise.
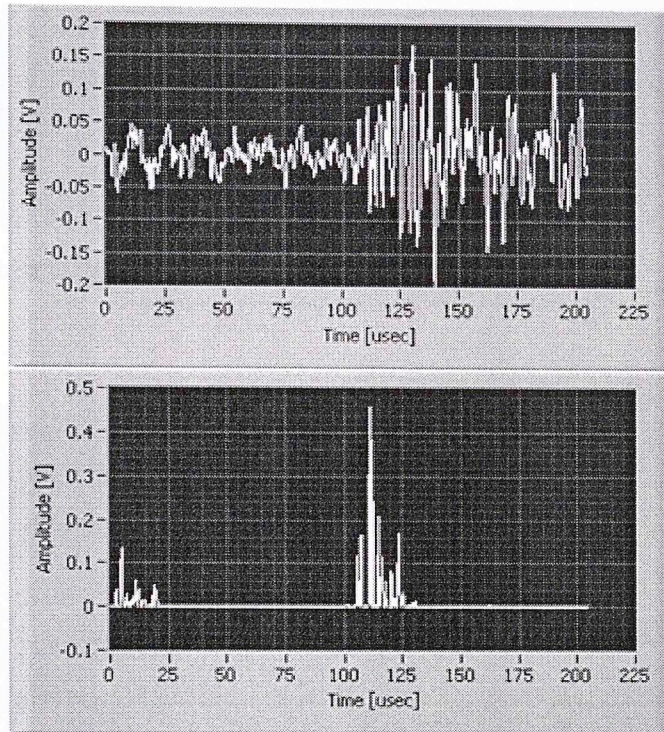
Figure 3.17 Modified energy ratio (MER) applied to a low SNR "medium" signal obtained from a real HF experiment. Window size 100 μsec.

Figure 3.18 shows the manual pick of an arrival on a "good" signal which has an SNR =28 dB, without any filtering applied. This signal will be used to analyze and compare the response of the CF to different MER window values. The results are presented on figures 3.19 to 3.25. The maximum peak of the MER characteristic function in the $er3_i$ ratio is closer to the manual arrival but there still is a small difference.

94
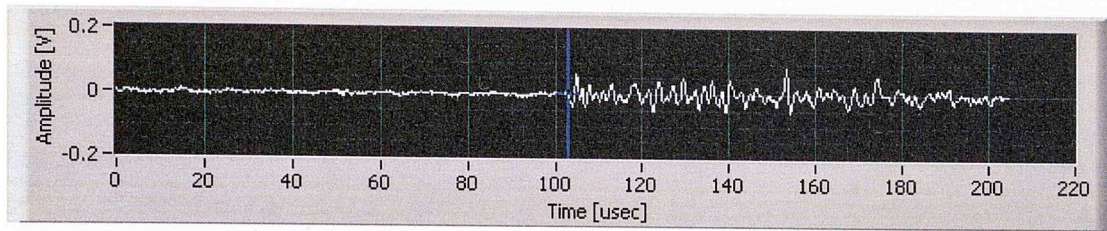
Figure 3.18 "Good" signal obtained from an hydraulic fracturing experiment. SNR = 28 dB. Manual pick obtained at 103.21 μsec.
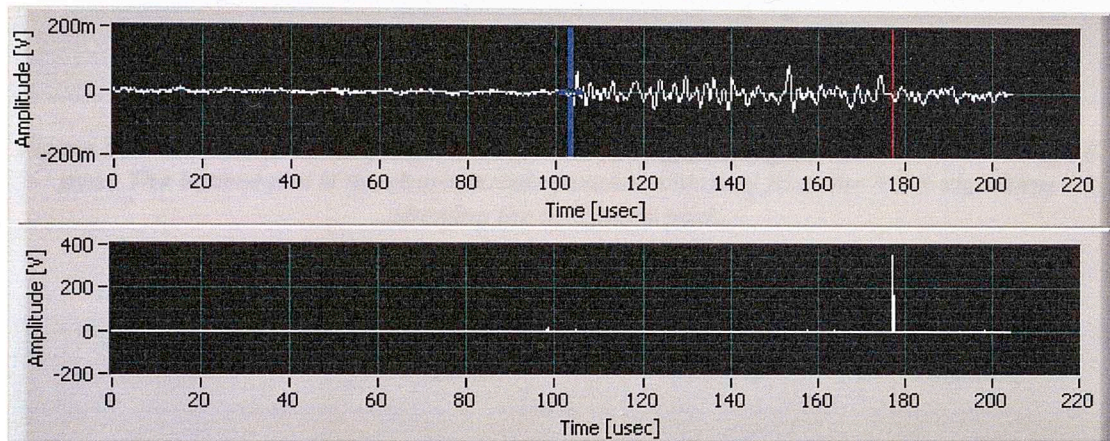
Figure 3.19 Manual (blue line) vs. automatic (red line) pick using an MER window size of 2 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.
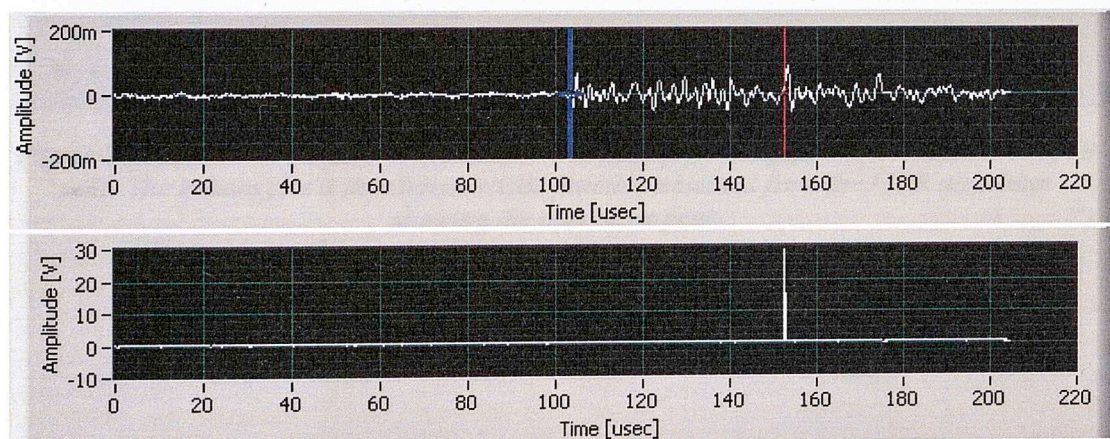
Figure 3.20 Manual (blue line) vs. automatic (red line) pick using an MER window size of 5 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.
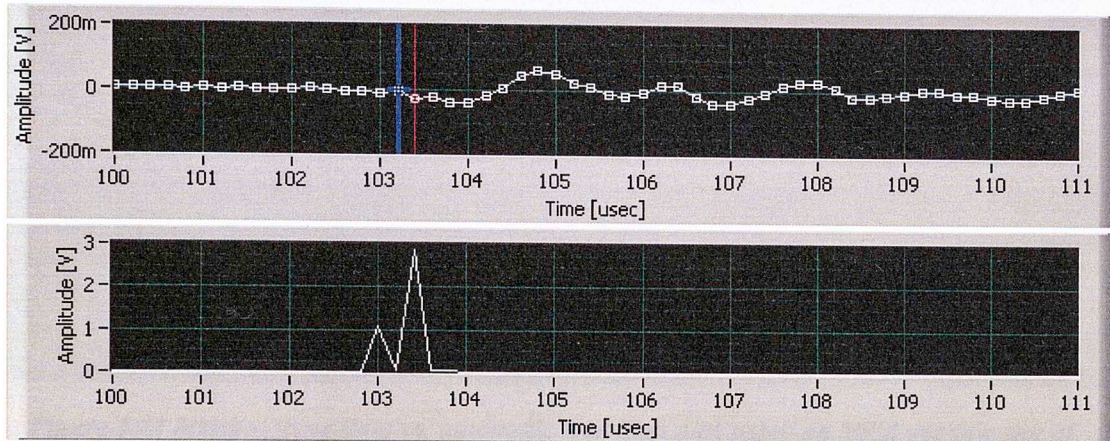
Figure 3.21 Manual (blue line) vs. automatic (red line) pick using an MER window size of 10 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.
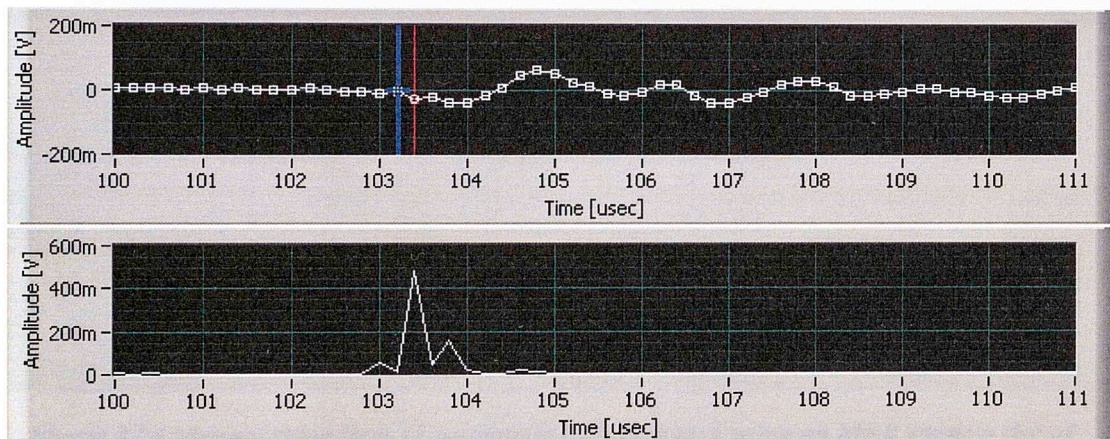
Figure 3.22 Manual (blue line) vs. automatic (red line) pick using an MER window size of 50 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.

96
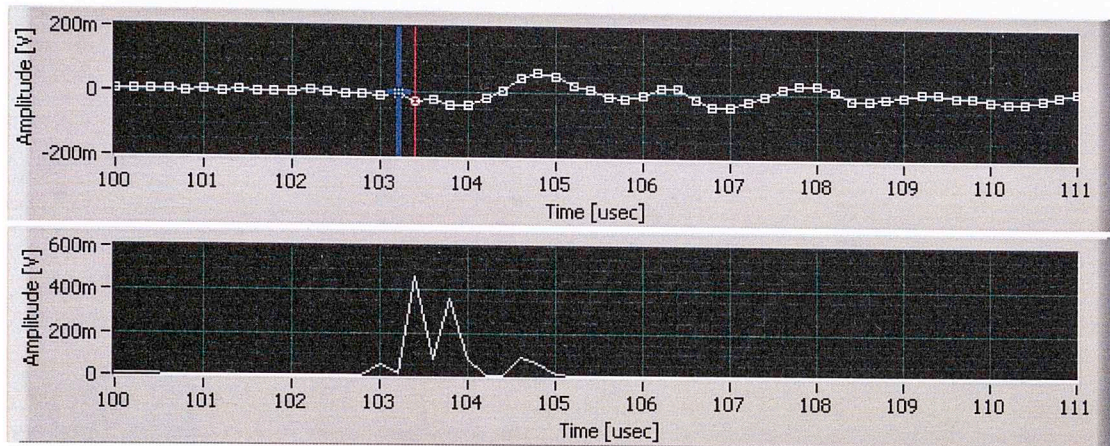
Figure 3.23 Manual (blue line) vs. automatic (red line) pick using an MER window size of 100 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.
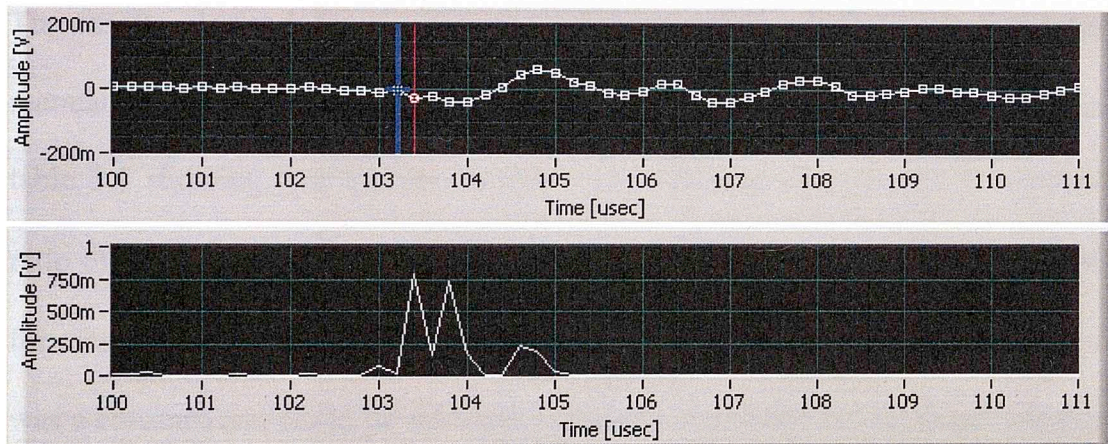
Figure 3.24 Manual (blue line) vs. automatic (red line) pick using an MER window size of 150 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.

97
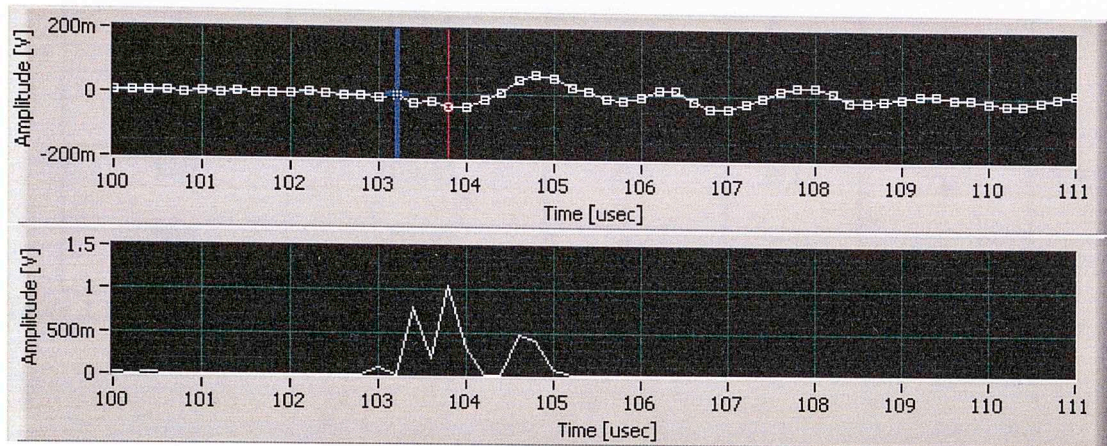
Figure 3.25 Manual (blue line) vs. automatic (red line) pick using an MER window size of 200 μsec. The bottom plot is the characteristic function obtained from the MER algorithm showing the maximum peak.
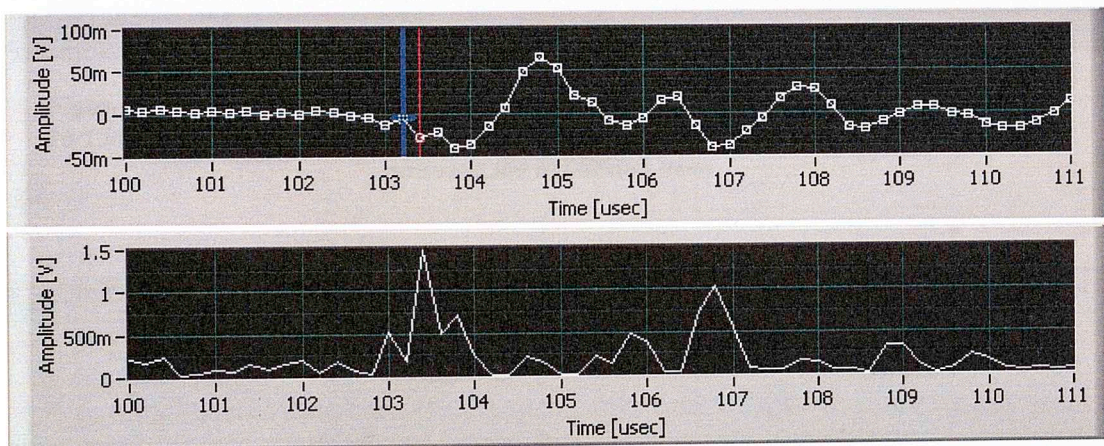
The results obtained for the different values of MER window ratios, and the manual comparison with the original signal in figure 3.18 are summarized in table 3.1 showing that the best results were obtained using an MER window ratio of 10, 50 and100 μsec with an error of 3 data points. This error is constant for most of the signals with a good SNR (SNR >10dB); for these cases this error was a constant and could be subtracted from the automatic arrival to get a better and accurate arrival pick.

MER Window [μsec]	Automatic Arrival Pick Time [μsec]	Absolute Error [μsec]	Relative Error %	Data points difference
2	177.2	73.99	71.69	370
5	152.6	49.39	47.85	246
10	103.4	0.2	0.18	1

50	103.4	0.2	0.18	1
100	103.41	0.2	0.18	1
150	103.41	0.2	0.18	1
200	103.81	0.6	0.58	3

Table 3.3 Error analysis using different MER windows. The manual arrival pick is measured at 102.81 μsec

The results using the MER algorithm show the efficiency in picking the arrival with a window size between 10 to 150 μsec with only 1 data point in difference (0.2 μsec) when compared with the manual pick.

This algorithm was also tested using synthetic signals with different SNR values. Arrival time was fixed at 60 μsec. Table 3.4 shows the result of MER automatic picking with a MER window set at 50 μsec.

Signal - SNR	Automatic Arrival Pick Time [μsec]	Absolute Error [μsec]	Relative Error %	Data points difference
(Inf) No Noise	60.20	0.2	0.33	1
73.9	60.20	0.2	0.33	1
63.2	60.20	0.2	0.33	1
54.4	60.20	0.2	0.33	1
42.8	60.21	0.21	0.35	1
33.24	60.25	0.25	0.42	1
23.44	60.30	0.3	0.5	1

11.9	60.25	0.25	0.42	1
8.56	60.44	0.44	0.73	2
6.45	60.47	0.47	0.78	2
3.44	60.39	0.39	0.65	2
1.34	24.20	35.8	59.66	179

Table 3.4 Analysis of error using MER characteristic function (50 μsec) with SNR variation on synthetic signals with an arrival set at 60 μsec. Note superior performance of this algorithm over the STA/LTA algorithm especially at lower SNR values.

The same synthetic signal used for STA/LTA algorithm with an SNR = 11.9 dB

is used now to show the benefits of MER algorithm in figure 3.26.
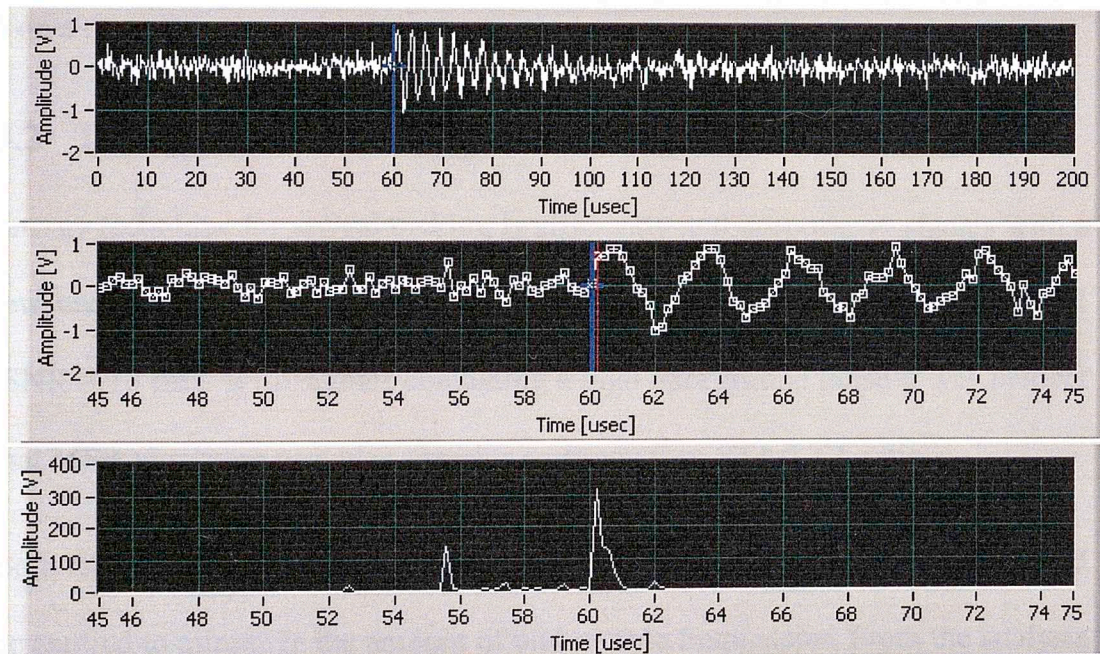
Figure 3.26 TOP: Original synthetic signal with arrival at 60 μsec (blue line). The following plot corresponds to a zoomed zone around the arrival with an automatic arrival detected at 60.2μsec (red line). The last plot (bottom) corresponds to the CF of the MER algorithm using a window of 50 μsec .

All results are extracted without any filtering applied and no onset time correction implemented. The time difference in between the manual arrival and the automatic picked arrival is only 1 to 2 data points. The MER algorithm performs better than the STA/LTA algorithm especially when the SNR is low, i.e. SNR < 10 dB. Table 3.4 shows that the MER algorithm was able to pick signals with a low SNR (3.44 dB).

Every algorithm will be need an onset time correction and MER is not an exception. MER will be tested on a complete set of signals obtained from a real HF experiment after classification. Complete results will be analyzed in chapter 5 on a complete set of classifications.

3.2   Test Results and Onset time correction

Energy technique for arrival picking is efficiently and accurate on signals with a low background noise, or, in other words, signals with a moderate high SNR (SNR > 10 dB). With signals containing a high background noise it was evident that MER algorithm provides superior accuracy than STA/LTA ratio.

Both techniques require a simple onset time correction. Also, a simple filter will be applied to minimize the amount of undesirable frequencies. From the analysis obtained on good signals collected and picked by Chitrala et Al., (2010) the filter is set to a bandwidth between 50 kHz to 1MHz using a simple Butterworth

type second order filter. Other filters and bandwidths are available for testing in the final software. (For analysis and settings on the use and the types of filters, please refer to LabVIEW Help.)

Analysis of different HF generated classified "good" signals using a combination of MER algorithm with a window of 100 μsec, filtering and time correction of only 3 data points results with the best accuracy possible. This combination of processing operations will be tested on the synthetic signals; the results are presented in table 3.5.

Signal - SNR	Automatic Arrival Pick Time [μsec]	Absolute Error [μsec]	Relative Error %	Data points difference
(Inf) No Noise	60	0	0	0
73.9	60	0	0	0
63.2	60	0	0	0
54.4	60	0	0	0
42.8	60.2	0.2	0.33	1
33.24	60.2	0.2	0.33	1
23.44	60.2	0.2	0.33	1
11.9	60.2	0.2	0.33	1
8.56	60.4	0.4	0.66	2
6.45	60.2	0.2	0.33	1

3.44	61.6	1.6	2.6	8
1.34	61.9	1.9	3.2	9.5

Table 3.5 Analysis of error using MER characteristic function (100 μsec), a 2ⁿᵈ order Butterworth band-pass filter (50 kHz- 1Mhz), and onset time correction with SNR variation on synthetic signals with an arrival set at 60 μsec.

Using this combination of bandwidth filtering and a MER window of 100 μsec the error is reduced for a signal with an SNR of 1.34 dB to only 3.2%. It is important to mention that the background noise created is random and this randomness even for a signal that has the same SNR could contain a different error. This will be analyzed for a complete test that combines classification and arrival picking for signals recorded during a real HF experiment in chapter 5.

4 LABORATORY MONITORING SYSTEM

AE/MS laboratory studies involve the detection and processing of events occurring in a finite body. In comparison to AE/MS field studies, data that is detected during laboratory studies will generally exhibit high dominant frequencies, lower signal amplitudes and signal complexities due to stress wave reflections from specimen boundaries (echoes or reflections). (Hardy, 2003).

4.1 Equipment and materials

This chapter explains and shows the entire experimental system configuration used to capture the AE signals during the hydraulic fracturing experiments. The complete system is depicted in the block diagram shown in figure 4.1.

The hydraulic fracturing system consists of a precision pumping system unit controlled by a personal computer (PC), a lateral stress system (flats jacks), an acoustic emission signal acquisition and processing system which is fed by a sensor or group of sensors, preamplifiers, a signal conditioning system and a Data Acquisition (DAQ) module attached to a PC and finally the test sample.

4.1.1 Pumping system

The pumping system generates the fluid pressure forcing the fluid down a borehole inside the rock which causes the hydraulic fractures when the fluid pressure overcomes the rock's tensile strength and the applied external pressure.

The pump used for this experiment is a Quizix SP-6200 pump system manufactured by Chandler Engineering and Ametek. The pump system is designed with two cylinders which provide continuous flow of the fracturing fluid. The system can operate in two different program modes; constant pressure or constant flow rate.

Figure 4.1 shows in detail the principal components of the system. Primary components are:

- 2 C-6000 Pump cylinders

- 2 constant volume valves

- 2 pressure transducers

- 1 CN-6000 pump controller

- Plumbing and cables

- Quixiz PumpWorks software operated through a PC

Two principal characteristics govern the operation mode of the pump. The first one is the flow mode, which must to be smooth and continuous using independent operation of both cylinders. This means that while one cylinder is delivering fluid the other retracts and pressurizes to the level of the other cylinder. Second is to deliver a constant flow rate.

The required pumping rates are different for different types of rocks; rates depend on porosity and permeability characteristic of each rock. The following table (see Table 4.1) shows the rate used for pumping fluid and some important characteristic information to the different kinds of rocks. Right circular cylinder rock specimens with diameters of 4" are used in this test.

	Indiana limestone (C)	Lyons sandstone (S)	Pyrophyllite (P)
Petrophysical characteristics			
Porosity, % Crushed Boyle's law @800 psi.	20 16 15	10 9 9	4 3 3
Permeability	5 md	20 μd	8 nd
Mineralogy, wgt%	Calcite, 95	Quartz, 85	
Sample and stimulation dimensions and conditions			
Length, in	5	4	4
Diameter, in	4	4	4
Borehole depth, in	2.5	2	2
Counter-bore depth, in	0.4	0.4	0.4
Perforation depth, in	2	1.6	1.6
Frac fluid viscosity, cp	Oil, 50	Oil, 50	Oil, 50
Pumping rate, cc/min	15	10	5

Table 4.1 Petrophysical properties and sample characteristics of rock samples under test. Taken from Chitrala et al. (2010).
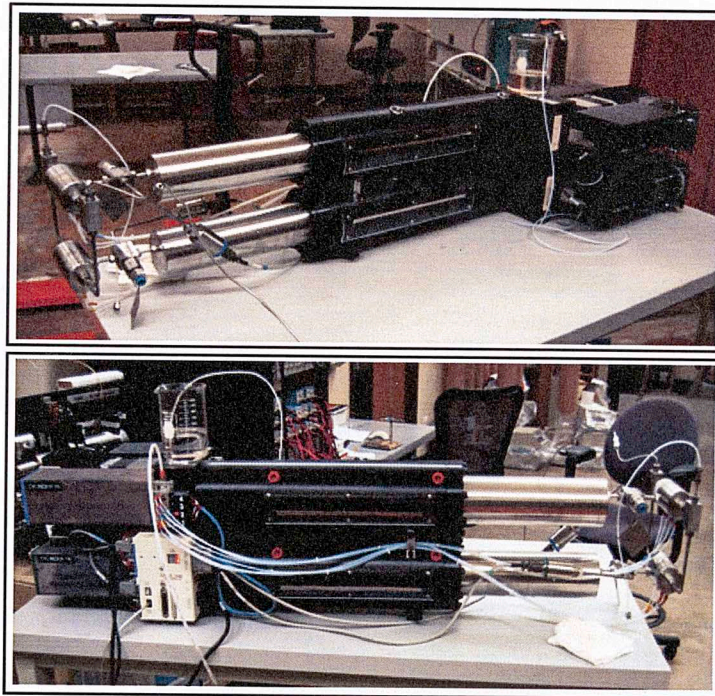
Figure 4.2 Dual cylinder pumping unit system. Top image is a front view of the system. Bottom image is the rear view.

The pumping system is connected to the rock sample by high pressure stainless steel tubing. The tubing connects to high pressure tubing which is epoxied in a hole bored into the sample.

4.1.2 Lateral stress system

Lateral stresses control the fracture direction. Applying a single horizontal stress causes a negative hoop stress at the azimuth of application and positive concentration 90 degrees away. Rocks are extremely weak in tension, thus the negative hoop stress controls the initiation of the hydraulic fracture. The

fractures are created by the differential stress between the lateral stress and the stress generated by the fracturing fluid inside the rock.
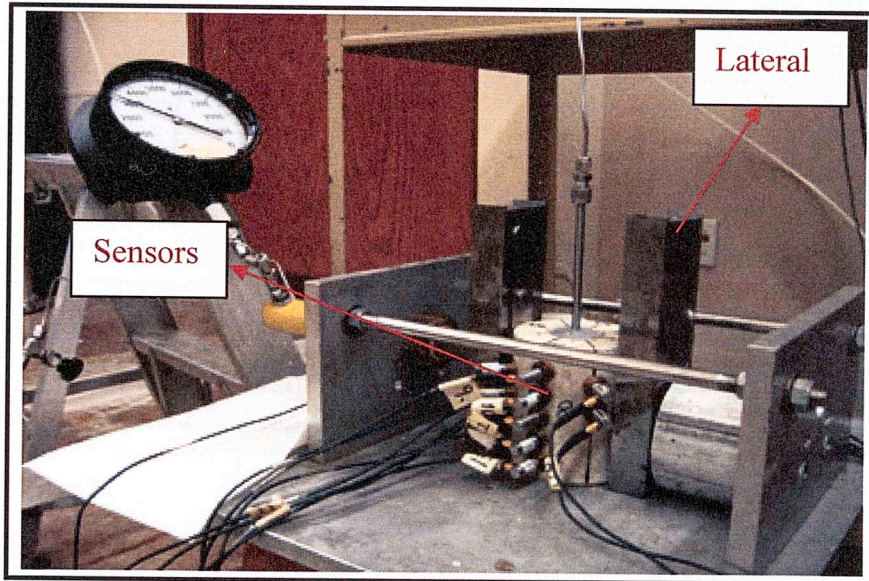
Figure 4.3 Position of rock sample on the lateral stress system. This picture shows the gauge for the pressure applied on the lateral sides of the rock. A sample is instrumented with 16 piezoelectric sensors distributed around and on top of the sample.

4.1.3   Fracturing Fluid

The two most common fracturing fluids are water and oil. For most of the experiments vegetable oil ($\mu = 50$ cp (centipoises)) was used.

4.1.4   Acoustic Emission monitoring system

This AE monitoring system is the most important part in the complete system, because it captures and stores the signals from the microfracture events. This system is capable of handling 32 sensors. The normal flow of the data across this system is illustrated on figure 4.4.

```
┌─────────────────────┐              ┌─────────────────────┐
│ 16-Channel Signal   │  ⇒           │ Data Acquisition    │  ⇒
│ Conditioning Unit   │              │ Module              │
└─────────────────────┘              └─────────────────────┘

    ⇑  Cables                                          ⇓

┌─────────────────────┐                            ┌─────────┐
│ Stack of 16 AE      │                            │         │
│ Pre-Amplifiers      │                            │         │
└─────────────────────┘                            └─────────┘
                                                    │   PC    │
    ⇑                                               └─────────┘
        16 AE Signals
        from Sensors
```
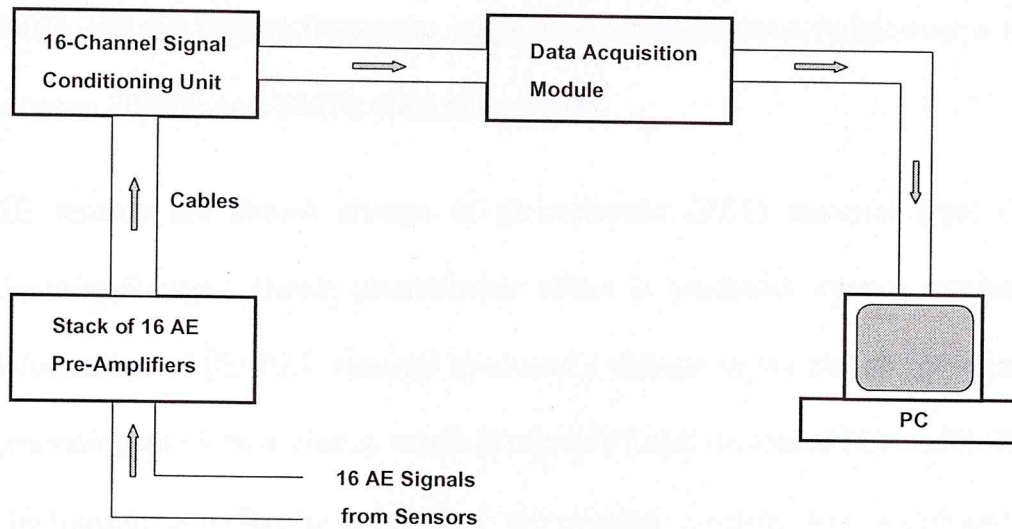
Figure 4.4. Block diagram of AE monitoring system (Aso, 2010). This figure shows the normal flow of data through the components of the AE monitoring system.

4.1.4.1 Sensors/Transducers

In general, sensors, also known as transducers, are piezoelectric elements that convert the acoustic energy into electric signals. The main purpose in AE experiments is to detect and capture the stress waves caused by local dynamic displacement and convert this displacement to an electrical signal (Muravin, 2008). AE activity at a specific point in the structure may be detected by monitoring the particle displacements, velocities or accelerations generated by the displacements at that point using a suitable transducer (Hardy, 2003).

Most laboratory studies in the geotechnical area involve the detection of AE events with dominant frequencies in the range of 50 kHz to 500 kHz (Hardy,

110

2003), but the typical frequency range in AE applications varies over a range between 20 kHz and 1 MHz (Muravin, 2008).

AE sensors are almost always of piezoelectric (PZT) material type, (lead-zirconite-titanate). Direct piezoelectric effect is produced when a mechanical deformation of the PZT material produces a change in the electric polarization generating an electric charge which is amplified and measured (Gautschi, 2002). The transduction element is usually a piezoelectric ceramic disk or cylinder with a thickness of a few mm. The thickness controls the resonant frequency of the transducer. An ideal transducer should be small, highly sensitive, easy to couple, cheap, easy to construct, with a high sensitivity over a wide frequency range (Hardy, 2003).

In this work, 16 B1025 piezoelectric transducers manufactured by Digital Wave Corporation (see figure 4.5) have been used. The B1025 are high fidelity piezoelectric transducers with a bandwidth of 50 kHz to 2 MHz. This sensor employs a low Q piezoelectric ceramic in a rugged package that is 0.365" in diameter by 0.50" in height. The principal attributes are a broadband flat response and general purpose testing (see Figure 4.6). Factory calibration suggests a good frequency response between 1 kHz to 1.5 MHz.

This sensor is specially designed for Modal Acoustic Emission (MAE), which is a method for determine the types of acoustic emission in materials using the

shape of the wave rather than just counting the events or using other parameters (Digital Wave anonymous). Each sensor is coupled to the test object by a thin layer of adhesive and its output is connected to a preamplifier via a BNC-to-Microdot coaxial cable (figure 4.5).

Figure 4.5. Broadband AE transducer. B1025 manufactured by Digital Wave Corporation.

These sensors are factory calibrated. A typical response is shown in figure 4.6. Sensitivity and frequency response are the principal criteria in the selection of these sensors. In general, for selection of sensors, the most sensitive sensor will be chosen commensurate with frequency response, size and availability. Sensitivity should be constant over the chosen frequency band (Scott, 1991).

Figure 4.6. B1025 Calibration curves provided by Digital Wave Corporation. Every sensor is also calibrated separately and provided by Digital Wave Corporation.

4.1.4.2 Couplant

The sensors are coupled to the rock sample using a mounting adhesive material (Crystalbond 555 Mounting Adhesive manufactured by SPI). This is a low melting point adhesive for moderate stress machining process with a viscosity at flow point of 500 cp which makes it perfect for temporally coupling the sensors.

4.1.4.3 Preamplifiers

The output of piezoelectric sensors is an electric charge. The magnitudes of the signals generated by AE waves are on the order of nV to mV (low level acoustic signals). The signals, which come from the sensor, in the preamplifier stage are filtered and amplified and then fed to the signal conditioning system.

The principal requirements for pre-amplification are bandwidth and gain. SNR and dynamic range are also taken into consideration. In this work, 16 Olympus Parametrics NDT 5660B wide band preamplifiers were used. They are 1MΩ input impedance, low noise, wide band (20 kHz to 2 MHz), with two manually selectable gains (40 dB or 60 dB).
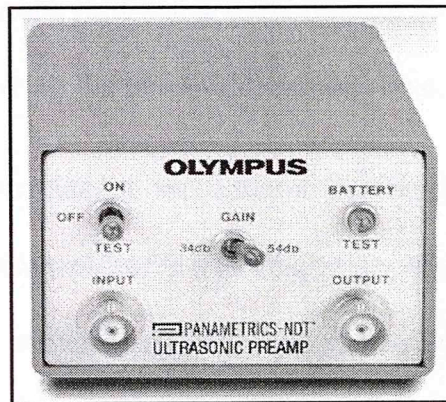
Figure 4.7. Olympus Parametrics-NDT. Preamplifier stage unit used for amplification of the signal coming from sensor.

Each preamplifier input is a BNC coaxial cable and its output a BNC coaxial cable going into the signal conditioning unit.

4.1.4.4 Signal Conditioning Unit

The preamplifier unit is used to amplify the signal; it is located as close as possible to the sensor in order to reduce noise. This step does not guarantee a completely clean signal. In order to reduce noise acquisition and preparation of the real AE signals a signal conditioning unit is used.

The FM-1 is a stand-alone signal conditioning unit manufactured by Digital Wave Corporation. This unit has 16 channels for conditioning signals in the frequency range from 20 kHz to 2.3 MHz and A/D converter boards. It consists of two controller cards which control the filter and gain settings for a group of 8 signal cards. Each signal card is used to condition signals from the preamplifiers. Every signal card has a BNC input with an impedance of 1 MΩ and receives the signal from the pre-amplification stage.

The controller card consists of an internal preamplifier, signal and trigger sections. The internal preamplifier is a low input impedance, low noise amplifier designed to work with a fixed gain high impedance preamplifier driver. It has thumb wheel switches providing a maximum of 42 dB gain in 6 dB increments. After amplification the signal is split into the signal conditioning section and trigger conditioning, this separation allows the system to prevent triggering on spurious noise whose frequency is outside the bandwidth of interest.
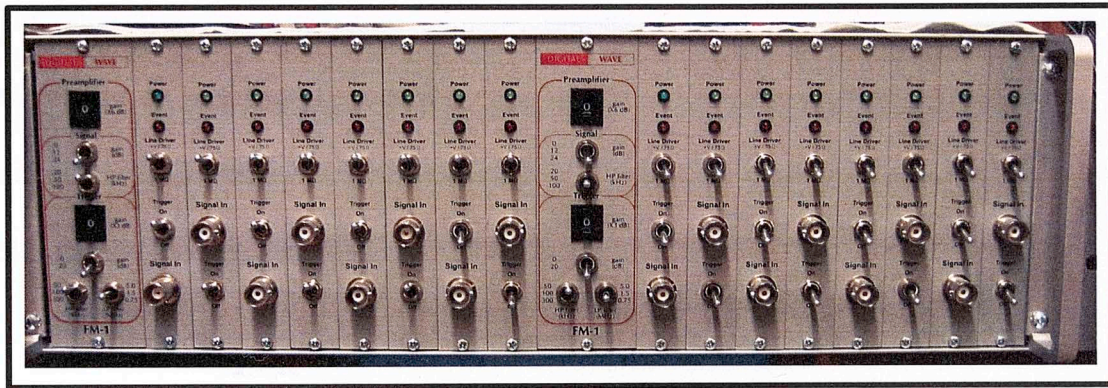
Figure 4.8. FM-1 Signal Conditioning Unit from Digital Wave Corporation. 16 individual channels.

The signal conditioning section has two options to modify the gain and filter of the signals. The gain toggle switch allows selection of an additional 24 dB gain in 12 dB increments. The high pass filter has a three position toggle switch to set between 20 kHz, 50 kHz or 100 kHz. The maximum signal frequency allowed is 5 MHz but the 3dB bandwidth at zero gain is 2.3 MHz.

The trigger stage conditioning can be adjusted in two separated gain and High Pass (HP) and Low Pass (LP) filtering options. It can give an additional gain of 21 dB adjustable in 3 dB increments and an additional gain of 20 dB. The filtering section of the trigger stage has the possibility to use a HP filter between 50 kHz, 100 kHz or 300 kHz while the LP filter can be set at 0.75 MHz, 1.5 MHz or 5 MHz. The trigger threshold is permanently set at 0.1 V by manufacturer.

To determine the total gain of the complete signal conditioning unit, it is necessary to add the preamplifier gain and the signal or the trigger gain. The signal, after it passes through the internal pre-amplification, is split and enters the signal or trigger stages.

STAGE		CONFIGURATION	
		GAIN [dB]	FILTER
Preamplifier	(External)	40 dB	Bandwidth (20kHz to 2 Mhz)
Signal Conditioning	Preamplifier (Internal)	0 dB	-
	Signal Stage	24 dB	HP 50 kHz
	Trigger Stage	0 dB	HP 50 kHz
		20 dB	LP 1.5 MHz
TOTALS		40+44=84 dB	Bandwidth 50 kHz - 1.5 MHz

Table 4.2 Configuration of Hardware stages on Acquisition of Acoustic Emission Signals.

4.1.4.5 A/D Board

The data acquisition module is the ICS-645 PCI Bus analog input board, specially designed for high frequency measurements. The principal characteristics of this A/D board can be summarized as follows:

- Thirty two 16-bit ADC channels sampling at up 20 MSPS (mega samples per second).

- 16 Mbytes of onboard storage

- PCI Bus (64-bit, 66MHz)

- Pre-trigger (circular buffer) data collection, in order to collect data both
  before and after the trigger.

The final step of the signal to the final visualization and storage is the analog to
digital conversion. A simplified block diagram for analog to digital converter is
shown on figure 4.9. For our purpose 16 channels were used.
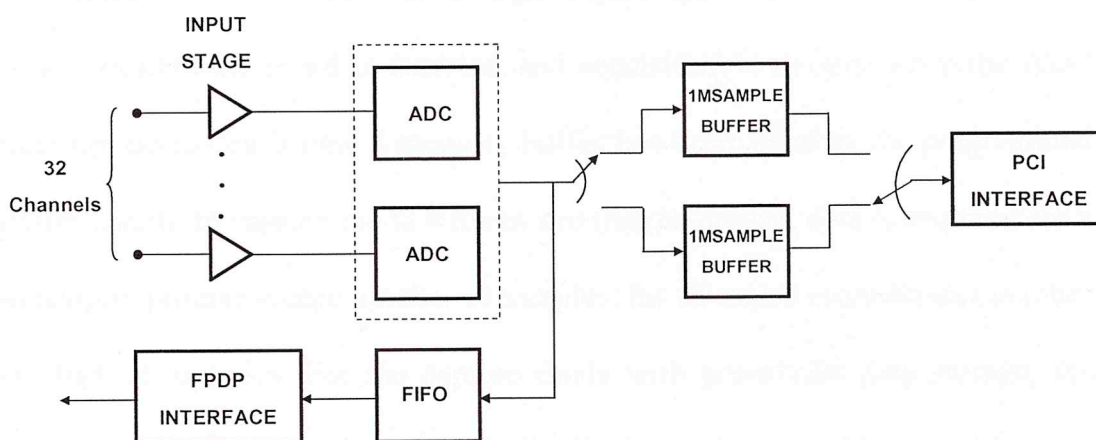
Figure 4.9 Simplified block diagram of the ICS-645. (Obtained from DaqScribe Technology, 2003). FIFO stands for First In Firs Out. FPDP stands for Front Panel Data Port Interface. ADC stands for Analog to Digital Converter. PCI stands for Peripheral Component Interconnect

The analog input stage of the board accepts single-ended input signals with peak
amplitudes of approximately ± 1.0V; the input impedance is 500 Ohms. A single
pole, low pass filter is included. This filter ensures a flat pass band response up
to the maximum bandwidth (5 MHz). The input stage is mainly composed by
four AD8056ARM voltage feedback amplifiers for amplification and filtering.

The output of the input stage is transmitted to the analog to digital converter (ADC) which employs an AD9260 chip per channel. Here, following the Nyquist criterion, the signals are sampled a rate which is at least twice the highest expected frequency in order to avoid aliasing.

Three principal operation modes are available for acquisition of signals. These are continuous mode and capture mode without pre-trigger and with pre-trigger. In continuous mode, acquisition begins upon application of the trigger and continues until the board is disabled, and acquisition is stopped when the ADC interrupt occurs each time a memory buffer has been filled to the programmed buffer length. In capture mode without pre-trigger storage, data is acquired for a maximum programmable number of samples; for all active channels this number is 1048576 samples. For the capture mode with pre-trigger data storage, the memory onboard is used as a circular buffer in a programmable circumference. The process is started by the user and the control logic continually fills the circular buffer with fresh data in anticipation of the trigger signal. Once the trigger signal is received, the final number of samples is stored in memory and acquisition is automatically terminated.
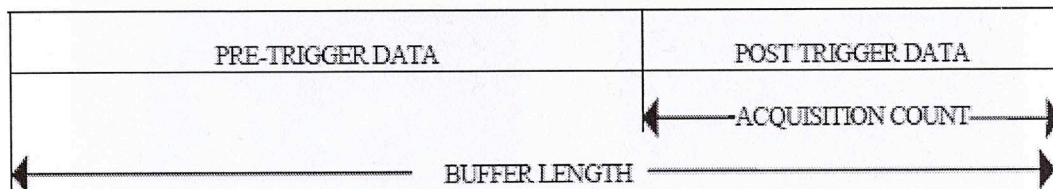
Figure 4.10 Operation of DAQ board on capture with pre-trigger mode. This graph shows the usage of the buffer memory for an acquisition.(Obtained from ICS sensor processing).

For our application, the capture mode is used with the pre-trigger mode. The total number of points is fixed on 1024 samples or 2048. This pre-trigger value signal is set by the user based on a percentage of the total number of samples. Normally this value is set between 10 to 50 %. Classification requires a pre-trigger setting that allows an analysis of the background noise and the signal. The normal value used is 50% on a 1024 samples signal, or 25% for a 2048 samples signal, figure 4.11 shows an example of an acquisition using different pre-trigger percentages. For an acquisition of a 10% the possibility of missing the arrival increases to a critical level, this is due to the threshold level of 100 mV. Some signals reach this level several samples after the first arrival; this behavior is normally seen on HF signals with low amplitude on the order of mV after amplification. Using a pre-trigger percentage of 25-50 % is sufficient to capture the arrival of the signal and ensure an amount of data to also sample background noise.
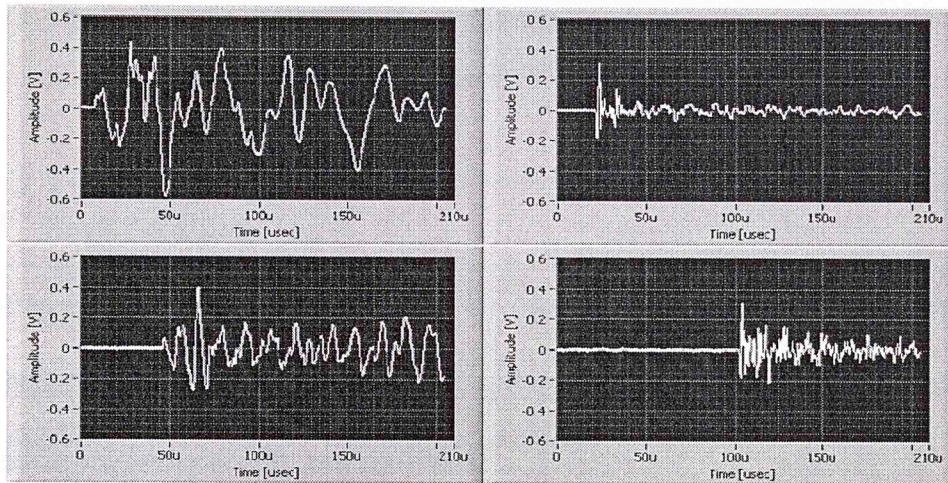
Figure 4.11. TOP LEFT CORNER: pre-trigger signal acquisition at 10%. TOP RIGHT CORNER: pre-trigger signal acquisition on 10%. BOTTOM LEFT CORNER: pre-trigger signal acquisition at 25%. BOTTOM RIGHT CORNER: pre-trigger signal acquisition at 50%.

Using pre-trigger percentages between 25 - 50 % and with a total number of samples of 1024, the number of samples used for noise analysis are between 256 - 512 samples or 51.2 - 102.4 μsec.

According to the Nyquist criterion, the sampling rate for data acquisition should be set so that the sampling frequency is at least twice the highest frequency expected. Hydraulic fracturing acoustic emission signals have a bandwidth range of 50 kHz to 1.5 MHz. The sampling rate should be at least 3 MHz. We used a 5 MHz sampling rate.

The sampling rate and the number of points together define the length of signal captured. For a 5 MHz sampling rate and 1024 points per channel, the time duration of the capture data is:

$$\frac{Number\ of\ points}{Sampling\ Rate} = \frac{1024\ points}{5x10^6\ points/second} = 204.8\ \mu sec$$

4.1.5 Rock Samples

Three principal rocks were used for HF experiments. These are listed on table 4.3 with a brief description on the principal characteristics that describes composition, permeability and porosity.

Rock Type	Mineralogy	Permeability	Porosity
Pyrophyllite (P)		80 nD	5%
Lyon Sandstone (S)	85% Quartz	20 μD	10%
Indiana Limestone (C)	95% Calcite	5 mD	20%

Table 4.3. Principal characteristics for the rock types used on the HF experiments. Permeability is measured in Darcy (D).

Rocks contain different types of grains and together with the voids and inclusions determine poroelastic properties of petroleum reservoirs. Fourier Transform Infrared (FTIR) Spectroscopy can be used in determining the mineral composition of rocks through comparison with a library of known mineral spectra (Aso, 2010; Sondergeld and Rai, 1993).

Permeability is measured in Darcys (D) and determines the ease with which a fluid can flow through a material. Darcy's law describes pressure driven laminar flow through porous media:

$$q = -\frac{KA}{\mu} * \frac{dP}{dx}$$

Where q is the fluid flow rate; K is the permeability of the medium; μ is the fluid viscosity; $\frac{dP}{dx}$ is the pressure gradient in the flow path within the medium.

Porosity is the ratio of pore to bulk volume and it is normally expressed as a fraction or percentage:

$$\phi = \frac{V_{pore}}{V_{bulk}}$$

Where V_{pore} and V_{bulk} are the pore and bulk volumes, respectively.

The dimensions of the samples are 4 inches in diameter and 3.5 – 4.72 inches in length. Some samples are shown in figure 4.12.
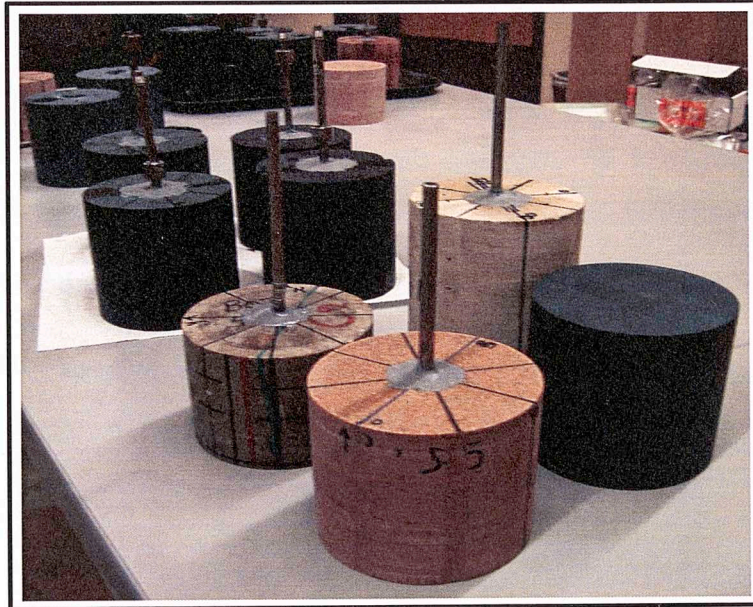
Figure 4.12. Rock samples prepared for HF testing. High pressure tubes are epoxied into a borehole and provide the path for the fracture fluid.

4.2   Experimental Procedure

Analyzing real signals obtained from rock samples is the final goal for this thesis. Chapters 2 and 3 show results obtained from different types of rocks; this chapter will explain the process to prepare the rock and obtain the final data (classification and arrival picking) from the large volume of experimentally generated events.

Three cylindrical core samples are obtained and carefully polished. A ¼ inch hole is drilled in one end and a counter-bore was made using a ½ inch coring bit and a steel mini-casing (0.12" internal diameter) with 2 perforations on both sides. One of the ends of the mini-casing has the perforation exit close to the

depth end and this end is cemented to the borehole wall using *Conley Weld* [TM]

epoxy (see figure 4.13).

To locate the sensors on the sample and to facilitate the installation of the mini-

casing, a Cartesian coordinate system (x,y,z) was used as reference. See figure

4.13 obtained from ASO, 2010.
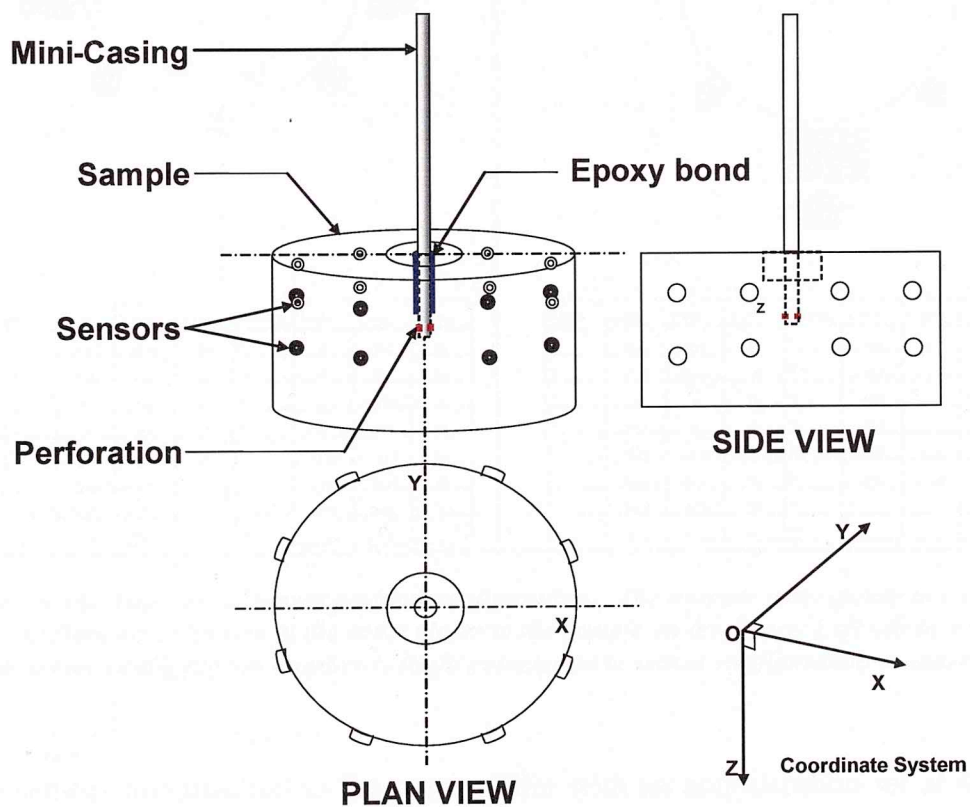
Figure 4.13. 3D plan and side view of a 4-inch diameter sample with a 0.12-inch internal diameter mini-casing. 16 AE sensors are shown surrounded the sample. Red squares indicate the perforation points while the blue portion symbolizes the epoxy glue that confines the fracture fluid around the perforation.

125

Sixteen AE transducers are attached to the sample using Crystalbond 555[TM] mounting adhesive. Figure 4.14 shows different configurations of the sensors.

Sensor #	X, mm	Y, mm	Z, mm	Event No	X, mm	Y, mm	Z, mm
1	-35.92	35.92	30	9	0	-50.8	20
2	-35.92	35.92	50	10	0	-50.8	40
3	0	50.8	30	11	0	-50.8	60
4	0	50.8	50	12	0	-50.8	80
5	35.92	35.92	30	13	-35.92	-35.92	30
6	35.92	35.92	50	14	-35.92	-35.92	50
7	35.92	-35.92	30	15	11.79	28.46	0
8	35.92	-35.92	50	16	-11.79	-28.46	0

Sensor #	X, mm	Y, mm	Z, mm	Event No	X, mm	Y, mm	Z, mm
1	35.92	35.92	30	9	-50.8	0	20
2	35.92	35.92	50	10	-50.8	0	40
3	50.8	0	30	11	-50.8	0	60
4	50.8	0	50	12	-50.8	0	80
5	35.92	-35.92	30	13	-35.92	35.92	30
6	35.92	-35.92	50	14	-35.92	35.92	50
7	-35.92	-35.92	30	15			
8	-35.92	-35.92	50	16			

Figure 4.14. Top view of sensor position configurations. The number corresponds to a sensor located from top to bottom in the order closer to the sample on the figure. Left will be known as sensor configuration number 1. Right correspond to sensor configuration number 2.

The sensors are attached to the preamplifier with an amplification set at 40 dB via a 6 ft long BNC to Microdot sensor cable.
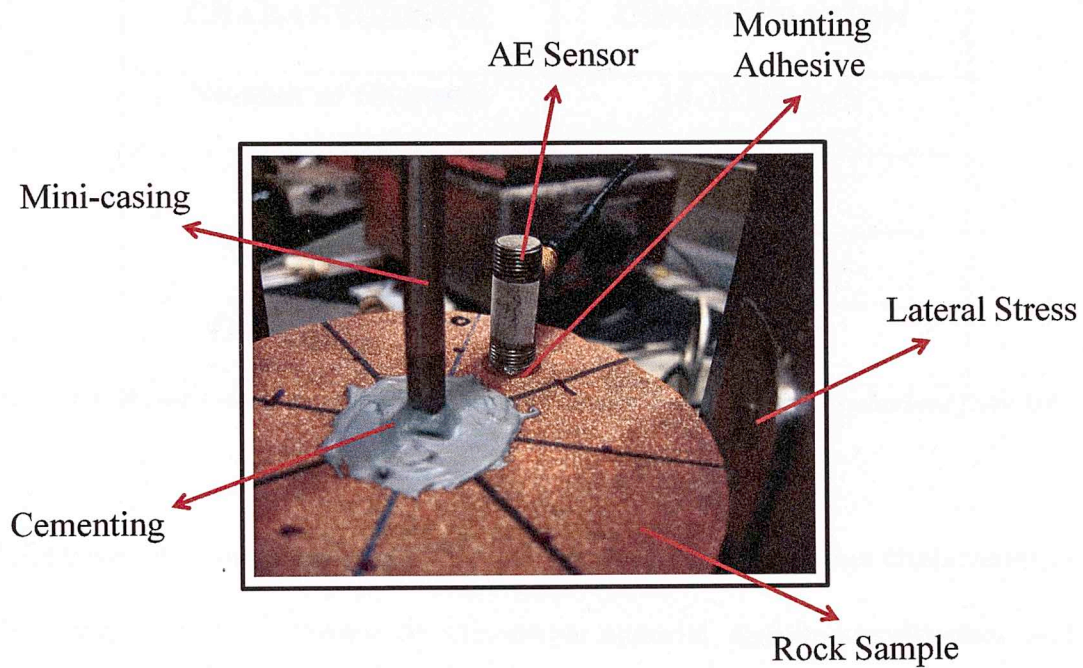
Figure 4.15 Sensor 16 attached to the top part of a sample by the mounting adhesive Crystalbond 555.

The signal conditioning unit receives the signal with the configuration given in table 4.4. The data acquisition is configured through the AE software (WaveExplorer version 7.1).

CHARACTERISTIC	CONFIGURATION
Number of Channels	14-16 channels
Sampling Rate	5 MHz
Number of Points	1024
Pre-trigger Points	25%-50%

Table 4.4. WaveExplorer configuration for data acquisition of AE signals obtained from HF experiments.

WaveExplorer requires location of every sensor to be input; other characteristics like velocity model, isotropic or anisotropic material, sample identification and elastic properties must be input too. A schematic of the complete process is showed in figure 4.16 which includes all the hardware involved on the system.

To ensure a good acquisition of the signals and a proper configuration of the hardware – software components, a calibration test is conducted. This calibration is done using the Hsu-Nielsen source calibration, using a pencil break (see figure 4.17) as a suitable simulated source of AE (Scott, 1991; Aso, 2010). The purpose of this exercise is to verify that the AE sensors are properly attached to the specimen, check transducer polarities, proved calibration data set for location algorithms and to provide a controlled data set for classification and arrival picking.

128
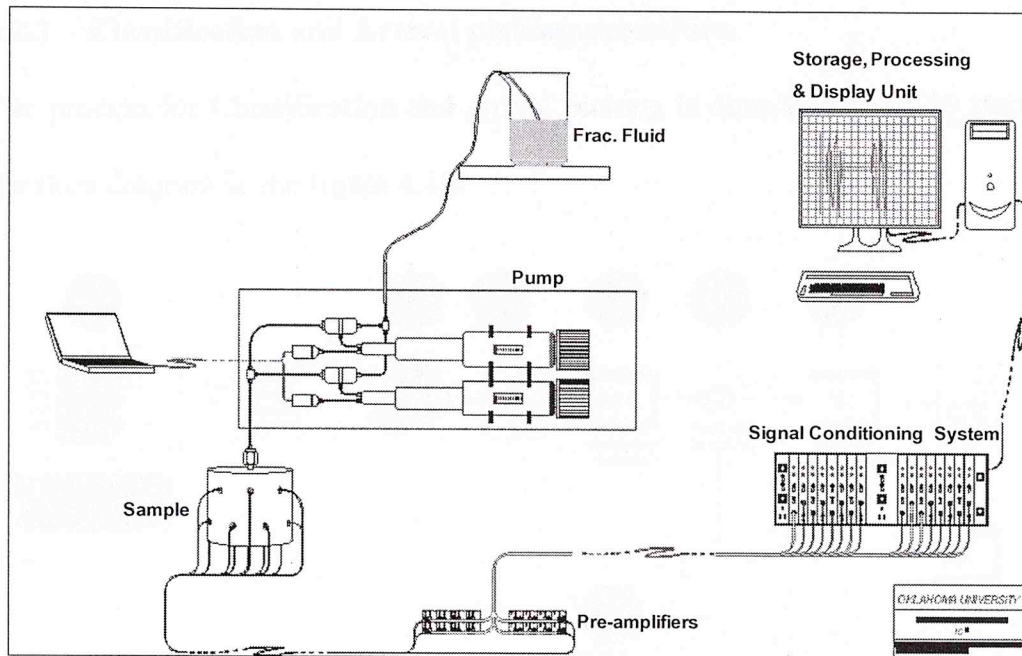
Figure 4.16 Schematics of Hydraulic Fracturing and microseismic monitoring system.
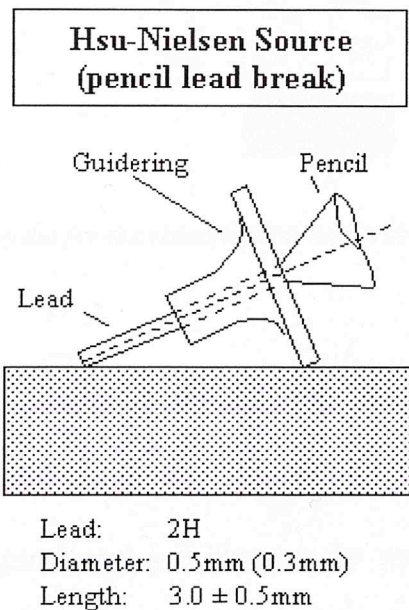
Figure 4.17 Hsu-Nielsen source for the testing and calibration of AE systems. (source: htpp://www.ndt.net/ndtaz/ndtaz.php).

4.2.1   Classification and Arrival picking procedure

The process for Classification and arrival picking is done following 10 steps of the flow diagram in the figure 4.18.
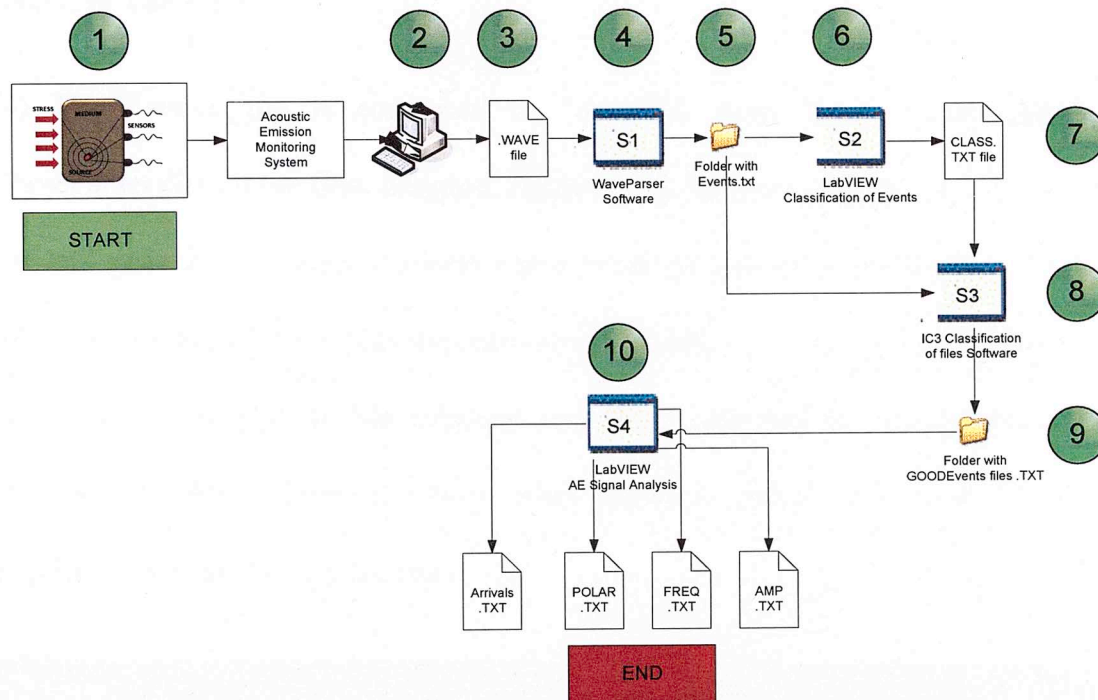
Figure 4.18 Flow diagram for the classification and arrival picking of AE events.

(1) The rock is prepared as mentioned before; the dimensions and the position of every sensor are entered into WaveExplorer for acquisition setup. (2) WaveExplorer is configured and a calibration by pencil-breaks is carried out. The events are recorded by WaveExplorer and are (3) saved as a *.WAVE file. This file contains all the events generated by the HF experiment and information about the acquisition: date, time, number of events, number of sensors, sampling

rate, etc. This file cannot be opened by any other software except from WaveExplorer, for this reason it has to be converted to a *.txt or *.lvm file. These are standard file extensions that are accessible to Microsoft Office (Word, Excel) or LabVIEW.

(4) The *.wave file is converted to *.txt file using the software called WaveParser (S1 on the flow diagram, figure 4.18). This program opens a *.wave file and generates a folder (5) with every event generated converted to a *.txt file. The number of *.txt files depends on the number of events captured in the *.wave file. Every *.txt file contains one event captured by the number of sensors used. The information from every sensor is saved as an array of the amplitudes sampled every 0.2 µsec.
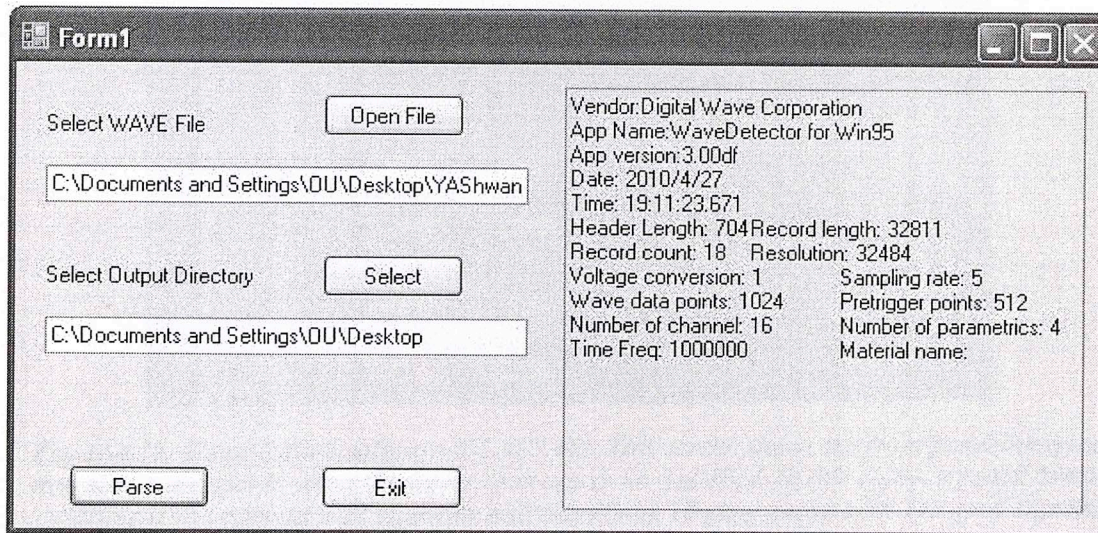
*Figure 4.19. Screen presentation of WaveParser (S1) software used to convert *.WAVE files to *.txt files. This image present the final information presented after a conversion of file extension. 16 channels for 18 complete events on a pencilbreak calibration experiment.*

Once the folder that contains every event with *.txt file extension is created, the files are readable by other software. LabVIEW can be used in this instance to make an automatic classification of the events. CLASSAE (6) is a LabVIEW program (located as S2 on the flow diagram, figure 4.18) which performs the automatic classification of AE events.

DIRECTORY

C:\Documents and Settings\OU\Desktop\
YAShwanth\P9-PB\P9-Pb breaks_event1.txt

SAVE ?

Event Number	Number of Sensors	Min Number of Good Signals	Number of BAD EVENTS	Number of GOOD EVENTS
18	16	6	8	10

Number of Signals
288

FINAL BAD INDEXED ARRAY

FINAL GOOD INDEXED ARRAY

0	0

FINAL BAD INDEXED ARRAY	FINAL GOOD INDEXED ARRAY
3	1
4	2
7	5
8	6
9	10
11	12
13	14
16	15
0	17
0	18

1) % Threshold   Range Value for Threshold 0.01

2) % Zero Crossings   Range Value for ZCross 0.03

3) % Histogram Middle Bin

4) Amplitude PtoP

5) Frequency Principal Component

6) Mean of Amplitudes   Start MPA 1 90   Window MPA 1 30

7) Mean of Frequencies   Start MoF 2 100u   Window MoF 2 100u

8) Signal To Noise Ratio   Start Noise 3 0   Start Signal 4 0.0001

Window Noise 3 0.0001   Window Signal 4 0.0002

Total Number of GOOD SENSORS
143

Figure 4.20. Classification software (CLASSAE). This screen shows the front panel interface that allows the user to select the folder of events to be classified. In this chase, a pencil-break experiment was selected with 18 events and classifying 10 good events with 143 good signals.

The input to CLASSAE is the location of the folder that contains the event files to be classify and the output (7) is a *.lvm (or *.txt) file (array type) containing the final classification information of the good events and good sensors. Table 4.5 shows an example of an output file resulting from the classification of 18 events (rows) from a calibration test of a lead-break using 16 sensors (columns). Every sensor of every event is analyzed, if a zero appears in the table this means that the sensor corresponding to the event was discarded.

E #	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	S 9	S 10	S 11	S 12	S 13	S 14	S 15	S 16
1	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
2	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
5	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
6	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
10	1	2	3	4	0	0	0	0	9	10	11	12	0	14	0	0
12	1	2	3	4	5	6	0	8	9	10	11	12	13	14	0	16
14	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
15	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16
17	1	2	3	4	5	6	0	8	9	10	11	12	13	14	0	16
18	1	2	3	4	5	6	0	8	9	10	11	12	13	14	15	16

Table 4.5. Classification results from a pencil break test. First column correspond to the event number. The other columns show the corresponding sensor of every event. If a zero appears on the spot corresponding to the sensor this means that the signal was not good enough.

This file can be opened by any software but normally is opened by default in Windows by Microsoft Excel to analyze the results obtained.

Once the classification file is constructed, it is necessary to create a new folder containing only the good events and the corresponding good sensors. For this reason the software called IC^3 Acoustic Event Detection (CLASSAE Eliminator) (referenced as S3 on the flow diagram) has been created (8). S3 receives the input folder containing all the events without classification (from S1) and the file output of CLASSAE (see table 4.5). The output of S3 is a folder containing every good event and every good sensor classified in a *.txt file (9). The number of files depends of the number of events classified as good events.

After the classification, the automatic arrival picking is performed on only "good" signals. Other wave analyses have been added to the software to enhance the information obtained from the signals. Polarization, frequency analysis, and maximum peak amplitude are important metrics computed by this software .This part of the signal analysis is done by S4 (10). S4 is the last step in the process; this software receives all the *.txt files already classified as "good" and applies the processes explained on chapter 2.

The software (shown on figure 4.21) performs and obtains the arrival picking, the polarization, frequency, and amplitude analysis. Each of these parameters are saved as an individual *.txt file ,ready for future analysis and reference for each experiment.
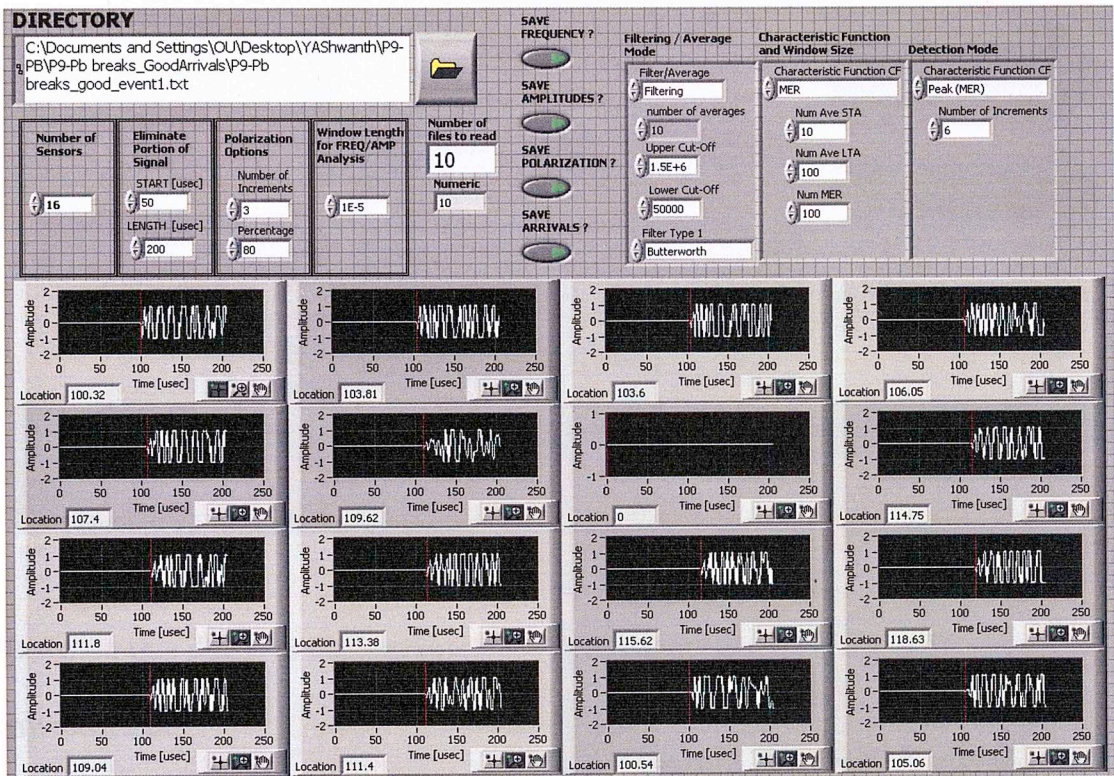
Figure 4.21 Front Panel of Signal Analysis for Microseismic Signals (arrival picking, polarization, amplitude and frequency analysis). Signals correspond to a pencilbreak test and shows the arrival time for every one during the complete process of analysis.

The process is straightforward trying to follow a normal and logical flow at each step. In general it is possible to recreate this process as a single box with only an input and multiple outputs as the one presented on figure 4.22.

135
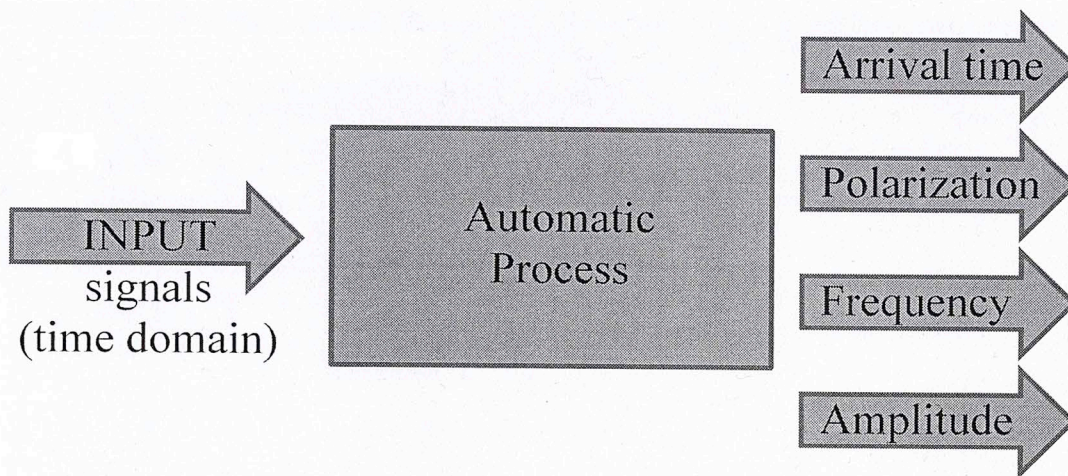
Figure 4.22 Single block diagram of the automatic process for AE signal analysis.

Based on this diagram the process flows in only one direction and does not need any feedback.

For a simple and complete analysis on a particular event or group of events, a third software called ONEbyONE (figure 4.23) has been created. This software lets you analyze every event for a particular experiment (contained on a folder). The analysis of results by events allows configuring different characteristic values to get an accurate arrival pick. Different visualization options also let a user analyze first arrival polarizations, arrival picks, amplitude and frequency.

Also, this visualization is recommended to be used before S4 software in order to review the classification results and to ensure a correct size of processing windows.
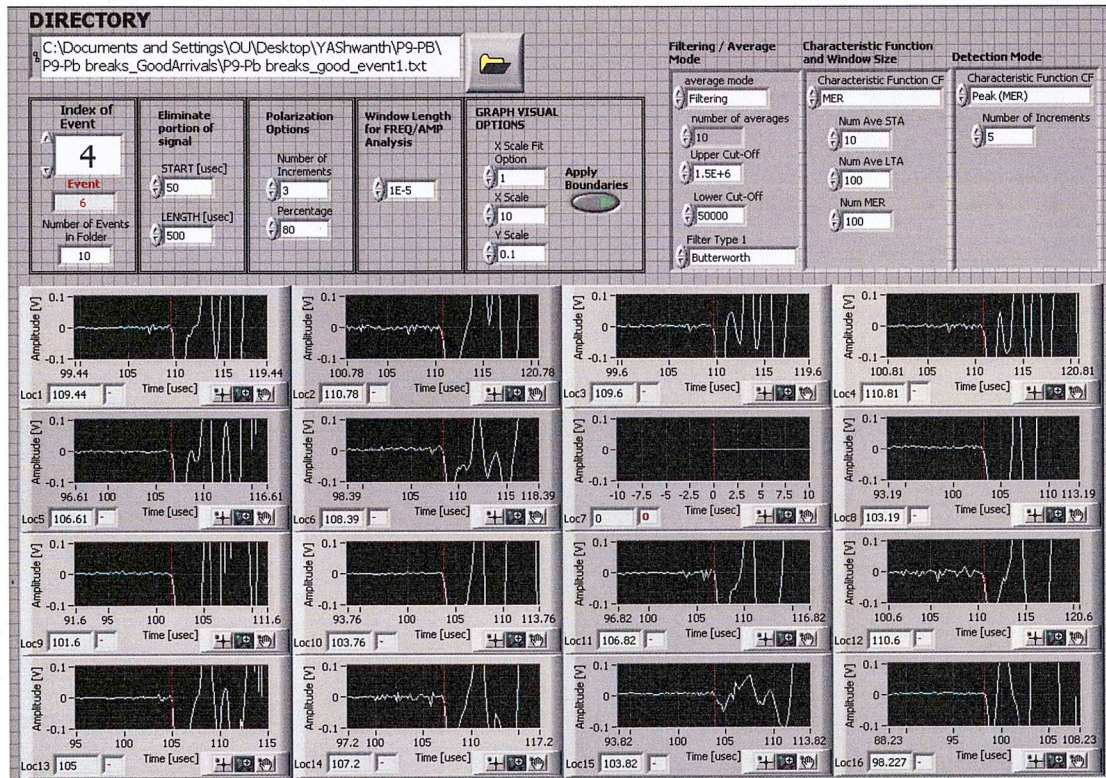
Figure 4.23 Screen display of page1 of OnebyOne software. Results from a pencilbreak calibration test showing a zoom around the arrival to ensure a right pick and a right analysis of the signal on every sensor. Sensor 8 corresponds to a bad sensor that was already eliminated by CLASSAE.

Visualization on a same plot (figure 4.24) is available in the second page of this software. This visualization permits analysis for every sensor arrival and shows the arrival automatically picked. Available on this page are the results from every event saved in the single *.txt file.
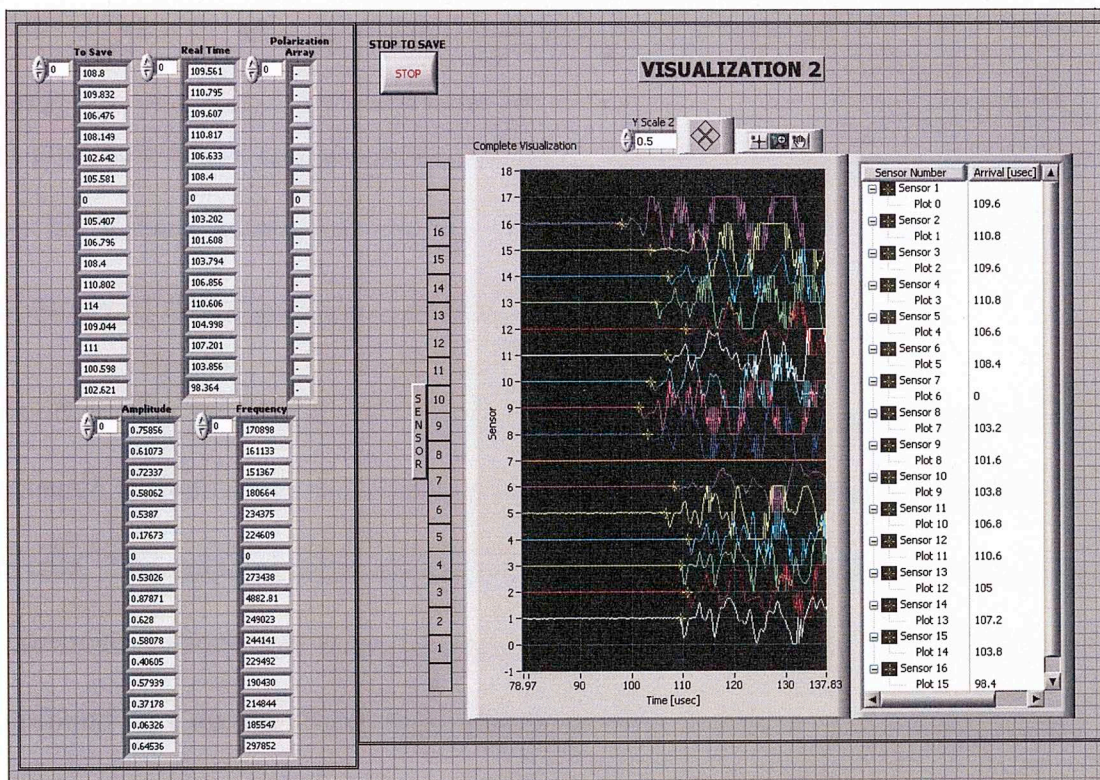
Figure 4.24 Screen display of page 2 of OnebyOne software. Final results from a particular event and automatic arrivals results on a same plot.

5   RESULTS AND DISCUSSION

A complete set of signals obtained from a real hydraulic fracturing experiment on an Indiana Limestone sample, C16, have been used to demonstrate the accuracy of classification and arrival picking.

Another way to measure the accuracy of the algorithms implemented in this research is to measure the impact or the improvement in the final location of AE events. The method used for final location will not be discussed here, but information about how to implement the final location based on the arrivals can be found on Castano, 2010; Stein and Wysession, 2003; Lay and Wallace, 1995 and Chitrala et al., 2010.

Locations of events allow visualizing where possible macro and microfractures occur within the sample. Hypocenters are determined through a minimization of a travel time inversion problem. The process starts with the end results, the seismograms or AE signals, and works backwards to minimize the calculated arrival times and ray paths consistent with the observed arrival times and sensor locations (Stein and Wysession, 2003).

5.1   Calibration Test

An initial calibration is carried out before the actual fracture experiment. This test will determine the accuracy for event locations. A pencil break is used to

simulate an AE event. Fracture of pencil lead causes an impulse event which takes places as the load on the surface is released (see figure 4.17).

In this test 8 pencil leads are broken in different previously determined surface locations. During the experiment some false events are also generated due to failure to break the lead or movements of the device around sample.

During the pencil break experiment on sample C16 a total of 20 events were generated. The results of manual classification of the signals are presented in table 5.1.

	Event number	Total number of Events
Good Events	1, 2, 4, 5, 6, 9, 10, 16, 18	9
Medium Events	3, 7, 11, 12, 19, 20, 15	7
Bad Events	8, 13, 14, 17	4

Table 5.1 Manual classification of a pencil break calibration signals on an Indiana Limestone rock sample.

Application of the different techniques explained on chapter 2 for classification result in the automatic classifications presented in table 5.2.

Classification Technique		Event number	Total number of Events	
T-M-S	Good	1, 2, 4, 5, 6, 9, 10, 16	Good	9
	Bad	3, 7, 8, 11, 12, 13, 14, 15, 17, 19, 20	Bad	11
T-H-Z	Good	1, 2, 3, 4, 5, 6, 7, 9, 10, 15, 16, 18, 19, 20	Good	14
	Bad	8, 11, 12, 13, 14, 17	Bad	6
T-M-S + 20%	Good	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 15, 16, 18, 20	Good	16
	Bad	8, 13, 14, 17	Bad	4

Table 5.2 Automatic classification results showing different algorithms implemented for a pencil break calibration test on an Indiana limestone sample.

Where, T corresponds to threshold algorithm; M corresponds to mean of frequencies; S to signal to noise ratio; H to histogram; Z to zero-crossing algorithms.

After reviewing the results from different algorithms for event classification, the algorithm that results in 100% of accuracy is the T-M-S (threshold – mean of frequencies - SNR). The classification technique implementing T-H-Z (threshold – histogram – zero crossing) results in an accuracy of 100% in detecting the good events but includes 5 "medium" events, which could lead errors on the arrival picking.

Finally a third algorithm was implemented trying to define the "medium" events. This technique will include the signals (and events) whose difference between the principal component 2 (PC2) and principal component 3(PC3) is less or equal to 20% of the complete maximum difference resulting from the principal component analysis (PCA). The results show that using TMS +20% include the "medium" events and exclude the "bad" events with 100% accuracy. This final technique will help to improve the results on signals with less diversity and could avoid possible "good" signals from the final classification. Diversity refers to the difference between signals and events acquired during a complete experiment. If the difference between "good" signals and "bad" signals is big the algorithms for classification could obtain better results (Tan, 2007).

The final results for location after classification using T-M-S algorithms for C16 pencil lead calibration are presented on figures 5.1. The method for arrival picking is MER algorithm (100 μsec window) and STA/LTA (10/100 μsec window) with a band-pass 2nd order Butterworth filter (50 kHz to 1 MHz). Figure 5.2 presents a comparison of event locations based on manual, MER and TSM algorithms.
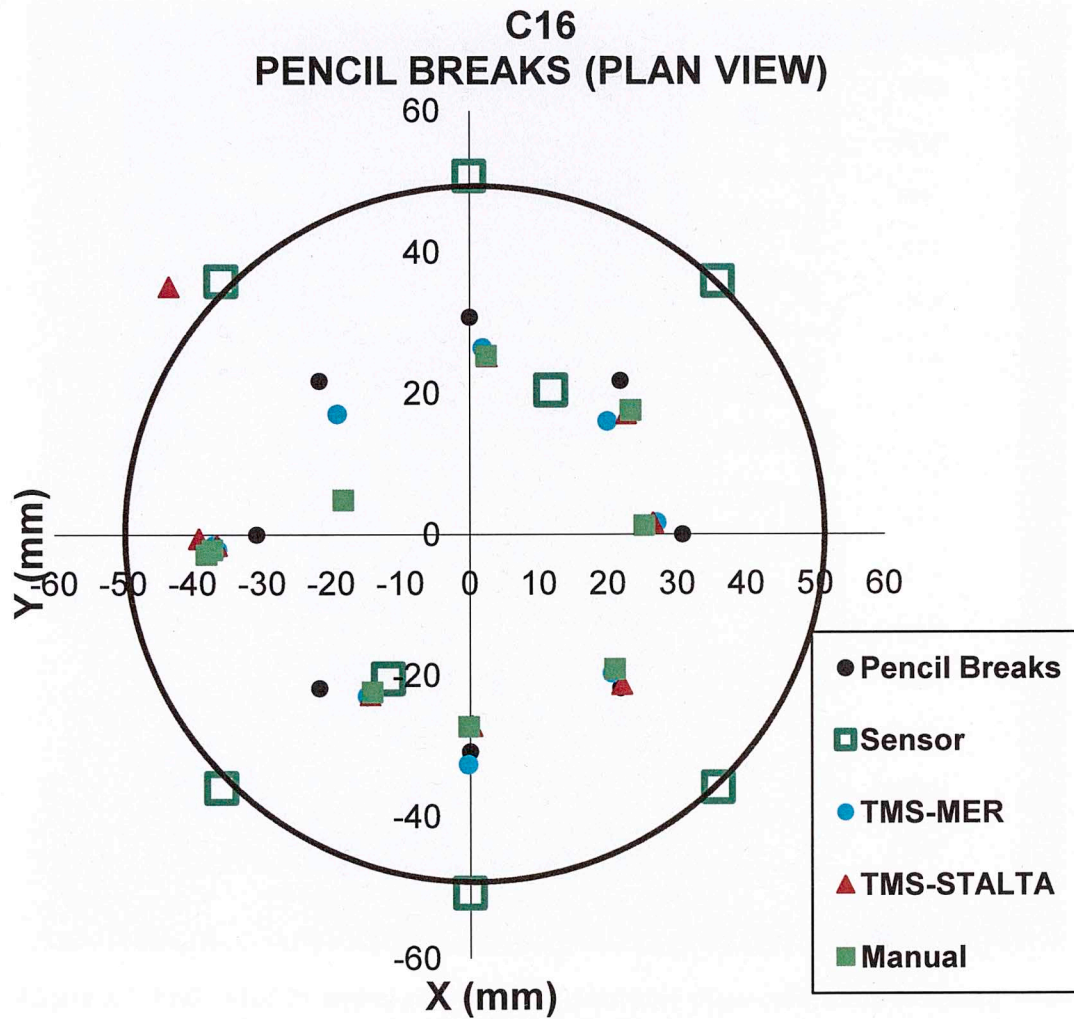
Figure 5.1 Pencil lead break calibration test location for events after manual and automatic classification (using T-M-S algorithms) and MER and STA/LTA arrival algorithms. T-M-S automatic classification and MER automatic arrival picking methods obtained the best results in comparison against manual methods.
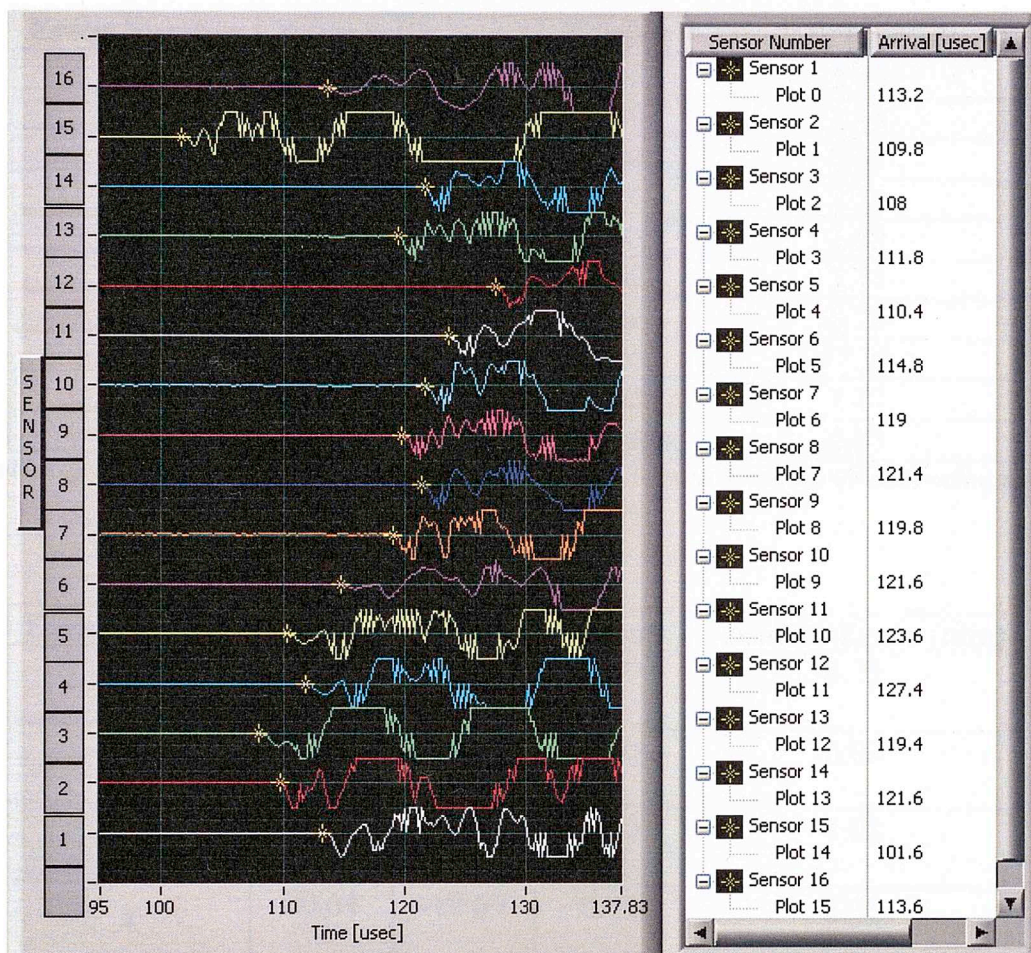
Figure 5.1 Event #1of 20 arrival picking for a pencil lead break calibration test using MER algorithm (100 μsec window).

Figure 5.1 plots the locations of events after automated classification and arrival time picking using T-M-S classification and MER (blue circle) and STA/LTA (red triangle). Also plotted are the hypocenters resulting from manual picking (green filled square). Results showing absolute error compared with the original location for manual picking are shown on table 5.3.

Event Number	X (mm)	Y (mm)	Z (mm)	Absolute Error (mm)
1	2.43	25.34	0.00	5.97
2	23.32	17.58	0.00	4.48
3	25.30	1.23	0.00	5.63
4	21.02	-19.14	0.00	2.74
5	-0.12	-27.28	0.00	3.52
6	-14.09	-22.37	0.00	7.71
7	-37.20	-2.18	0.00	6.76
8	-18.25	4.86	0.00	17.28
9	-38.03	-2.85	0.00	7.77
			Average	6.87

Table 5.3 Manual RMS error hypocenter location for C16 pencil lead break calibration after manual classification.

Event Number	X (mm)	Y (mm)	Z (mm)	Absolute Error (mm)
1	1.92	26.43	0.00	4.77
2	19.95	16.05	0.00	6.02
3	27.08	1.66	0.00	4.08
4	20.64	-19.69	0.00	2.38
5	-0.24	-32.62	0.00	1.84
6	-14.79	-22.94	0.00	7.08
7	-36.95	-1.58	0.00	6.35
8	-36.46	-2.18	0.00	6.07
9	-19.13	17.09	0.00	5.38
			Average	4.88

Table 5.4 Automatic classification (TMS) and arrival picking (MER) absolute error hypocenter location for C16 pencil lead break calibration.

Event Number	X (mm)	Y (mm)	Z (mm)	Absolute Error (mm)
1	2.52	25.26	0.00	6.08
2	22.68	17.11	0.00	4.75
3	26.54	1.59	0.00	4.55
4	22.02	-21.26	0.00	0.57
5	0.34	-27.17	0.00	3.65
6	-14.41	-22.76	0.00	7.43
7	-36.54	-1.55	0.00	5.95
8	-39.14	-0.61	0.00	8.36
9	-43.46	35.11	0.00	25.45
			Average	7.42

Table 5.5 Automatic classification (TMS) and arrival picking (STA/LTA) RMS error hypocenter location for C16 pencil lead break calibration.

The average absolute error for location for hypocenter locations from manual classification and manual arrival picking is 6.87 mm. Using automatic (TMS) classification and MER automatic picking is 4.88 mm absolute error. Finally using automatic classification and STA/LTA automatic picking the result in average absolute error is 7.42 mm. Absolute errors are calculated using the following formula:

$$Absolute\ error = \sqrt[2]{(X_c - X_r)^2 + (Y_c - Y_r)^2 + (Z_c - Z_r)^2}$$

Where X, Y, Z are the axis locations, c correspond to the value calculated and r correspond to the real value of location.

After reviewing results for locations using manual and automatic classification and arrival picking, it is clear that the best results were obtained using automatic classification and MER automatic arrival picking with an absolute error location of 4.88 mm and 0.4% (0.004 μsec) absolute error in time comparison vs. manual results.

5.2   Hydraulic Fracturing Test

A complete test for HF on the same Indiana limestone sample following the procedure mentioned in chapter 4 was carried out. An entire dataset of 269 events using 16 sensors generates 4304 signals. These were automatically classified and each arrival time was picked automatically. The results for manual and automatic classification are summarized in table 5.5. Due to the number of signals, the events were manually classified simply as "good" or "bad" events. Figure 5.3 shows an example of a good event (event 44) selected randomly to show the results obtained for every signal (of every sensor) using the MER automatic arrival picking algorithm.

Classification Technique	Total number of Events	
	Good	Bad
T-M-S	68	201
T-M-S +20%	269	0
T-M-S +2%	204	65
T-S-H	136	133
T-S-H + 20%	210	59
T-H-Z	149	120
T-H-Z + 20%	210	59

Manual Classification	206	63

Table 5.6 Automatic and manual classification results on events for an Indiana limestone HF experiment. Total number of events was 269, and 4304 individual waveforms were analyzed.

The time implemented for a manual classification and arrival picking of the total set of events is approximately 8 hours versus approximately15 minutes to automatically classify and pick arrival times.
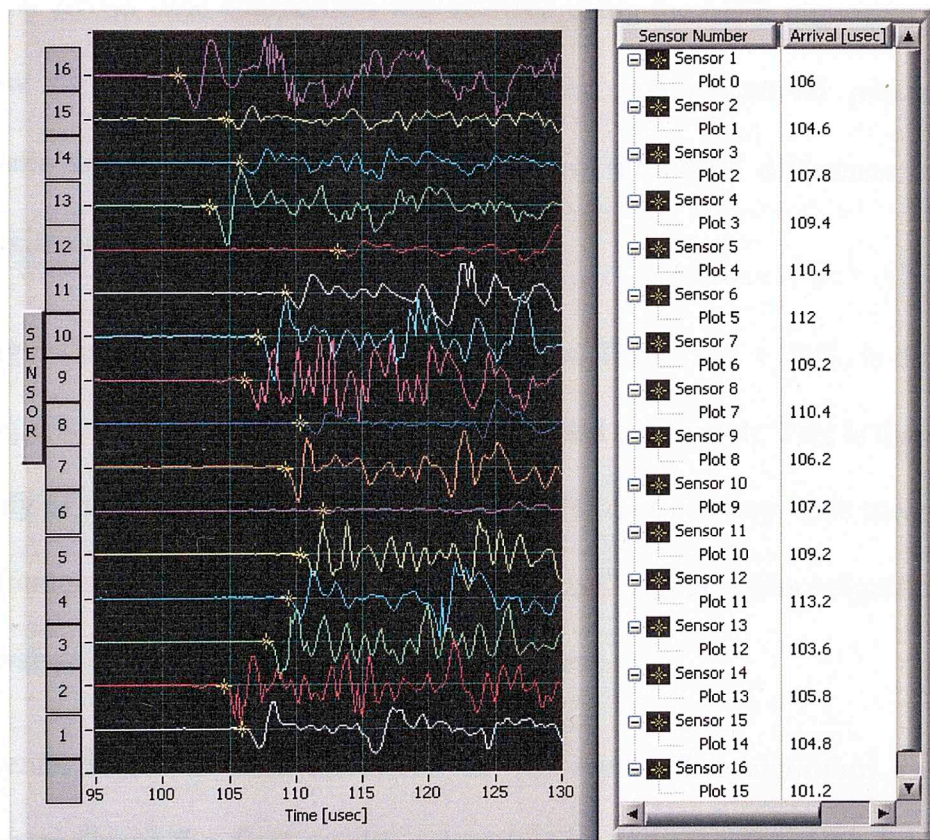
Figure 5.2 Automatic arrival picking results (yellow cursor) using MER algorithm (window = 100 μsec) for a "good" event number 44 of 269. The different signals correspond to every sensor (16 in total). Sensor 1 plotted at the bottom to sensor 16 plotted in the top. On the right of the plot, the time results in μsec for automatic arrival are also shown for every sensor.

The dependency of automatic classification depends on the diversity of signals analyzed. This is something also mentioned by Tan, 2007. Table 5.5 shows a summary of results obtained from implementing different automatic classification algorithms. The best results are obtained using threshold – signal to noise ratio – histogram (T-S-H) + 20% which includes "medium" class signals with a total number of 210 events classified as "good", creating 98.1 % (4 events more declared as "good") of accuracy compared with manual results.

The best arrival time comparison was realized using the MER algorithm with a window of 200 μsec. The results comparing manual arrival picking vs. automatic arrival picking give an absolute error of 1.99% difference in time (approximately 1.93 μsec in time).

The algorithm implemented for HF experiments, The T-S-H + 20%, is different in comparison with the T-M-S used for pencil lead calibration. This is due to the diversity of the signals which in the case of an HF test is lesser. This means that the difference between a "good" signal and a "bad" signal is more significant in a calibration test than in a hydraulic fracturing test.

Hypocenter locations were calculated for HF events. The algorithm used for location produces the final results generating an error measured by the RMS (root medium square) error. This algorithm depends on a velocity model obtained by other process for the rock. After a manual classification and arrival

picking is made and the location algorithm is applied, the uncertainties in arrival times will produce mislocation of events, sometimes outside the sample. This signals located outside the sample are eliminated.

Table 5.7 shows the root medium square (RMS) error for location for different classification and arrival picking methods. The RMS error is calculated from the error estimates generated during the least squares event location. A comparison is presented based on the automatic and manual picks.

METHOD		RMS error (mm)	Number of events located inside sample (Total number of events)
Classification	Arrival Picking		
T-M-S (65 "good" events)	MER(100μsec)	3.97	65
	MER (200μsec)	0.50	65
	STA/LTA(10/100μsec)	2.05	61
T-S-H (136 "good" events)	MER(100μsec)	3.46	131
	MER(200μsec)	0.97	128
	STA/LTA(10/100μsec)	2.89	119
T-S-H+20% (210 "good" events)	MER(100μsec)	2.76	172
	MER(200μsec)	1.71	177
	STA/LTA(10/100μsec)	2.85	165

Manual (210 "good" events)	Manual	0.57	201

Table 5.7 Analysis of RMS error obtained after location of events inside sample using the methods that yields on best results.

The automatic classification method that produced the largest number of "good" events (177 "good" events) and a good RMS error (1.71 mm) is using threshold – signal to noise ratio – histogram plus 20% which includes "medium" class events.

It is important to clarify the relation between classification of "good" events and accuracy on final location. In order to achieve small RMS error in locations (< 1 mm) it is necessary to eliminate "medium" level signals. This level of accuracy is reached by the single T-S-H classification method which classifies 128 "good" events with an RMS error in location (using MER 200 μsec) of 0.97 mm and eliminates 74 "medium" events.

Some examples on location of events from HF on Indiana limestone sample are presented on figure 5.4, 5.5 and 5.6 using different axis views (X-Y; X-Z and Y-Z respectively).
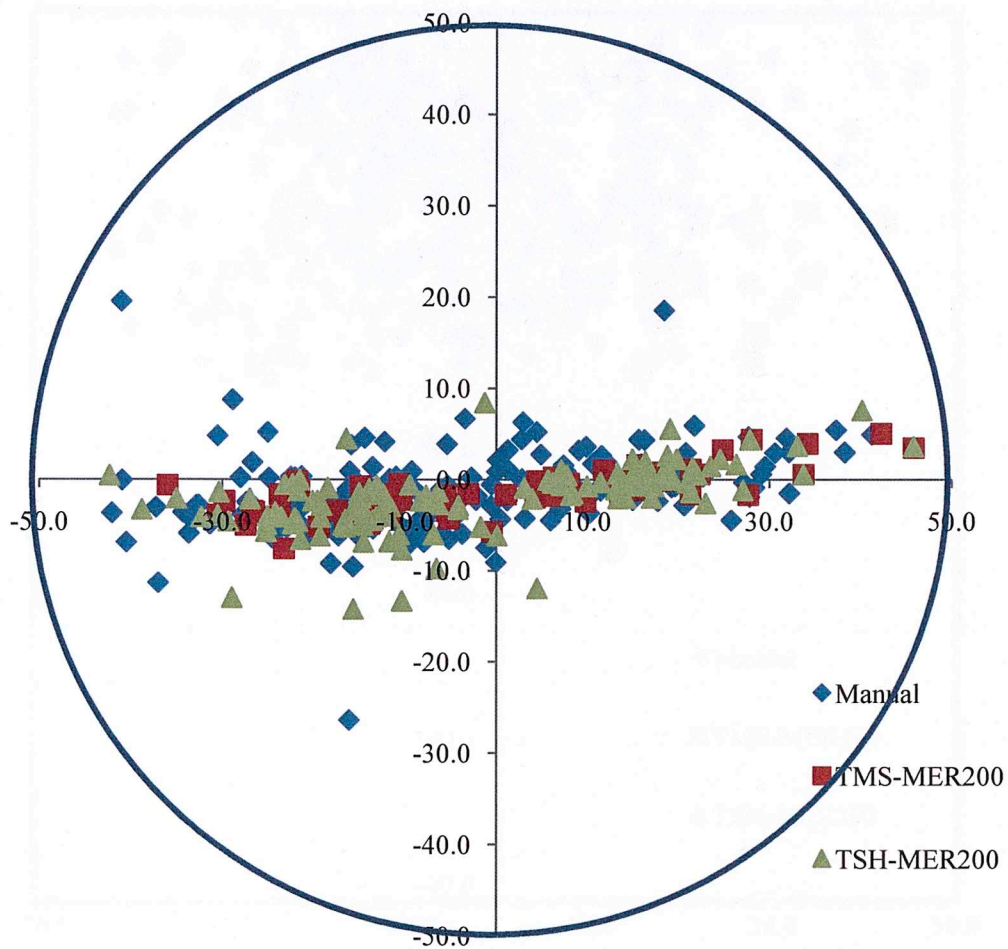
Figure 5.3 Event location example using the classification and arrival picking methods that yields best results. X-Y axis plan view; distance in mm. Blue squares correspond to manual methods for classification and arrival picking. Red squares correspond to T-M-S automatic classification method. Green triangles correspond to T-S-H automatic classification method. Both automatic methods for classification were finally located using MER automatic arrival picking algorithm with 200 μsec window.
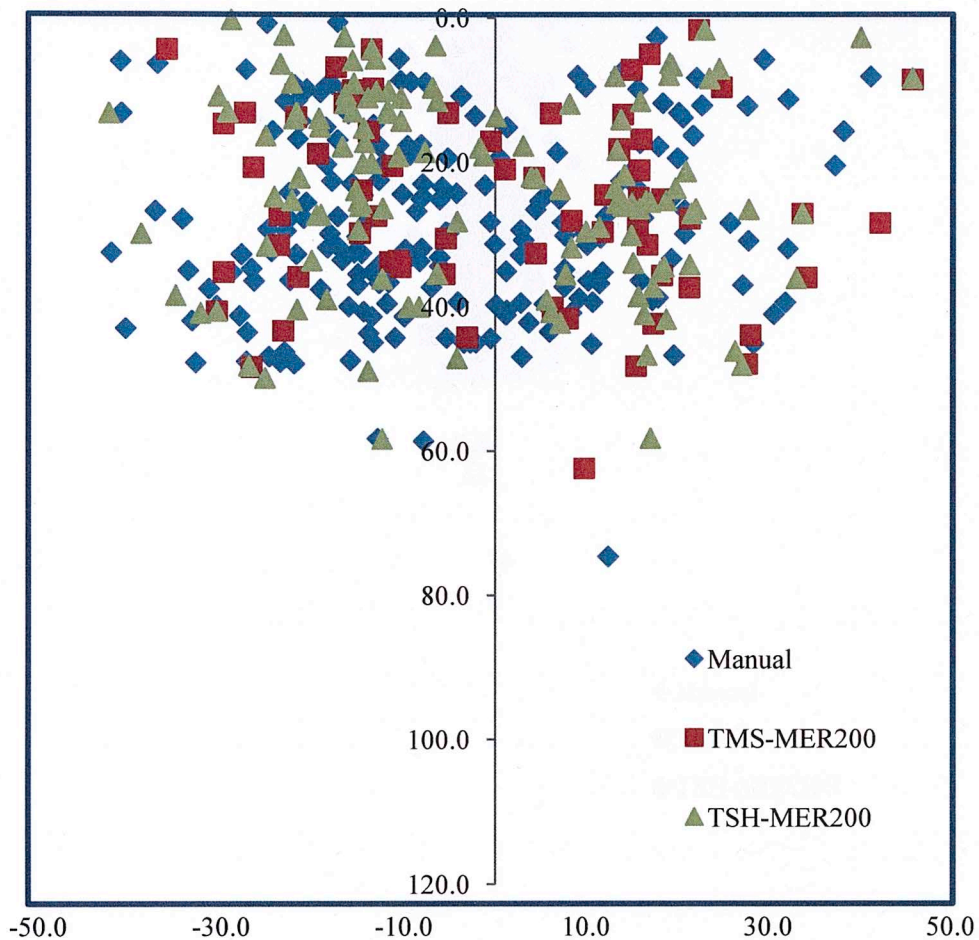
Figure 5.4 Event location example using the classification and arrival picking methods that yields best results (; distance in mm). X-Z lateral axis view projected onto the diametrical plane of the sample. Blue squares correspond to manual methods for classification and arrival picking. Red squares correspond to T-M-S automatic classification method. Green triangles correspond to T-S-H automatic classification method. Both automatic methods for classification were finally located using MER automatic arrival picking algorithm with 200 μsec window.

154
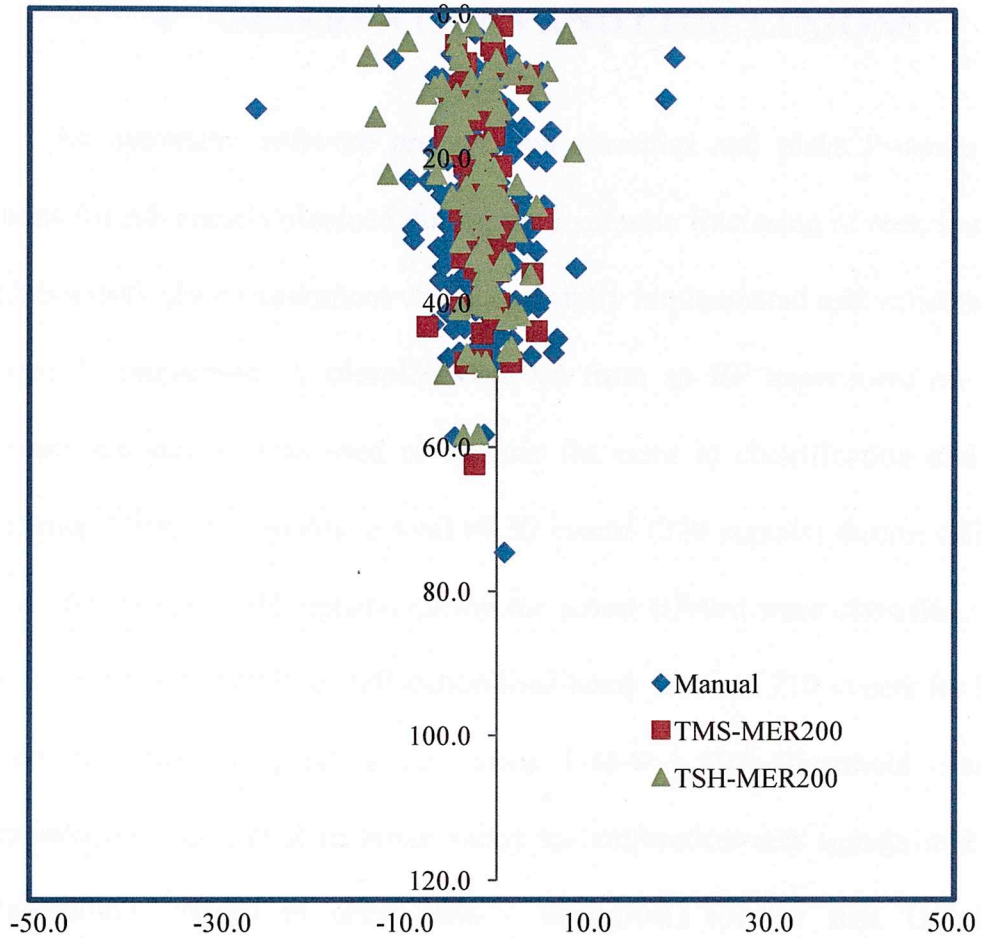
Figure 5.6   Event location example using the classification and arrival picking methods that yields best results (; distance in mm). Y-Z lateral axis view projected onto a diametrical plane perpendicular to the final hydraulic fracture. Note the tight clustering of events in the region of the final fracture. Blue squares correspond to manual methods for classification and arrival picking. Red squares correspond to T-M-S automatic classification method. Green triangles correspond to T-S-H automatic classification method. Both automatic methods for classification were finally located using MER automatic arrival picking algorithm with 200 μsec window.

6 OBSERVATIONS AND CONCLUSIONS

An automatic software process that classifies and picks P-waves arrival times for AE signals obtained during the hydraulic fracturing of rock samples in a laboratory scale experiment was successfully implemented and validated using manual comparison. A complete data set from an HF experiment on Indiana Limestone sample was used to analyze the error in classification and arrival picking. Using 16 sensors, a total of 20 events (320 signals) during calibration and 269 events (4304 signals) during the actual HF test were classified. Results indicated that 9 events in calibration lead-break test and 210 events for HF test were classified as "good" events using T-M-S + 20% (threshold – mean of frequencies and signal to noise ratio) for calibration test signals and T-S-H (threshold – signal to noise ratio – histogram) for HF test. Overall this combination of algorithms produces a 98.1 % agreement in classification with manual classification. Other algorithms yield the following levels of classification agreement: 72% using T-H-Z, 66% using T-S-H and 33% using T-M-S.

Results for automatic arrival picking (MER and STA/LTA) algorithms were also compared against manual results. MER arrival picking algorithm gives the best results with a 0.4% of absolute error in time comparison vs. manual arrival

156

picking in pencil lead break calibration and 1.99% of error in HF real test signals.

Final event locations are also examined to demonstrate the consequences in the final interpretation of AE data. The best results for classification were obtained using a combination of 3 algorithms (threshold – mean of frequencies – signal to noise ratio) including "medium" signals. Inclusion of "medium" signals will increase the error in the automatic arrival picking method in about 1 or to 2 decimals in RMS error. It is recommendable not to use "medium" signals in order to improve the uncertainty and RMS error location.

Two principal programs, one for classification and another for arrival picking have been created for processing AE events from laboratory experiments. A third program lets the user interact with properties and algorithms used for arrival picking and the characteristics at the individual signal level. This third software could be used to determine which combination of characteristics performs the best. Different algorithms were analyzed and used for both classification and arrival picking and are available for analysis on the selectable options of the software.

Additional signal information is extracted from signals after classification; these characteristics are: first arrival polarization (positive, negative or not possible detection), principal frequency around a window

centered on the arrival and analysis of maximum amplitude in this window. All three use the arrival time to define where to obtain those characteristics and depend on a good performance of the arrival picking procedure. These signal characteristics are important for the characterization and analysis of AE events results (Chitrala et al., 2010).

The results show that the time required for automatic picking is shorter than any manual method. Automatic pickers have improved to the point where they can now rival human analysts' results on large volumes of data. The time implemented on average for a set of 269 events in manual classification and manual arrival picking is approximately 8 hours compared to 15 minutes on average to finish the complete automatic processes.

MER algorithm method developed by Wong et al., (2009) is simpler and faster than the STA/LTA window ratio approach. Both methods are effective at picking arrivals on low noise seismograms (SNR < 10dB). But as the random noise increases, both methods begin to fail. The method with better accuracy and greater noise tolerance is the MER algorithm, whose results and performance are discussed in chapter 5.

Future work should look to implement an "online" automatic procedure. The automatic procedure system implemented in this research works "offline." This means that the classification and arrival picking is implemented after the

acquisition is done. The acquisition of the signals is implemented by a different

system. It is possible for future implementation to realize a complete acquisition

and analysis on "real time." This is extremely important in the field where it is

necessary to study fracture stages as they are carried out. This will require new

software which is directly coupled with the acquisition software.

REFERENCES

Alderson, F. (2004). Toward a three-dimensional crustal structure of the Dead Sea region from local earthquake tomography. PhD thesis. University of Tel-Aviv.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, Bulletin of the seismological society of America, 68, 1521-1532.

Allen, R. V. (1982). Automatic phase pickers: Their present use and future prospects. Bulletin of the Seismological Society of America, 72, 6, p.S225-S242.

Aso, I. I. (2010). Microseismic mapping of Hydraulic Fractures, MS Thesis, The University of Oklahoma

Bai, C. Y., and Kennet, L. N. (2000). Automatic phase-detection and identification by full use of a single Three-component broadband seismogram, Bulletin of the seismological society of America, 90, 1, pp. 187-198.

Castano, A. (2010). Estimation of uncertainty in microseismic event location associated with hydraulic fracturing, MS Thesis, The University of Oklahoma.

Chitrala, Y., Moreno, C., Sondergeld, C., and Ray, C. (2010). Microseismic mapping of laboratory induced hydraulic fractures in anisotropic reservoirs. Mewborne School of Petroleum and Geological Engineering, The University of Oklahoma, SPE 138441.

Digital Wave Corporation. (2010). Modal Acoustic Emission (MAE), Equipmentand testing services. Retrieved from http://www.digitalwavecorp.com/modalae-dwc.htm

Douglas, A. (1997). Band-pass filtering to reduce noise on seismograms: Is there a better way?, Bulletin of the seismological society of America, 87, 4, p.770-777.

Grosse, C. U., and Ohtsu M. (2008): Acoustic Emission Testing, Basic for Research – Applications in Civil Engineering, Springer.

Han, L., Wong, J., Bancroft, J. C., and Stewart R. R. (2009). Automatic time picking and velocity determination on full-waveform sonic logs, 2009 CSEG Conference.

Hardy, R. H. (2003). Acoustic emission / microseismic activity, Vol.1, Principles, Techniques and Geotechnical Applications.

Jackson, J. E. (1991): A User's Guide to Principal Component Analysis., John Wiley and Sons.

Lavrov, A. V., and Shkuratnik, V. L. (2005). Deformation and fracture induced acoustic emission in rocks. Acoustical Physics, Vol. 51, Suppl. 1, pp. S2-S11.

Leonard, M. (2000). Comparison of manual and automatic onset time picking. Bull. Seism. Soc. Am., 90, 6, pp. 1384-1390.

Leonard, M. (2000): Comparison of Manual and Automatic Onset Time Picking, Bulletin of the seismological society of America. 90, 6, pp. 1384-1390.

Lindsay, I.S. (2002): A tutorial on principal component analysis, Student Tutorial, University of Otago, New Zealand.

Montgomery, D.C., Runger G.C., and Hubele N.F., (2001): Engineering Statistics, Second edition, Arizona State University.

Munro, K. (2004): Automatic event detection and picking of P-wave arrivals. CREWES research report.

Munro, K.A., (2005): Analysis of microseismic event picking with applications to landslide and oilfield monitoring settings, University of Calgary. Calgary, Alberta

Muravin B., (2008): Acoustic Emission Basic, Acoustic Emission Method, History, Fundamentals and applications. www.muravin.com

Scott, I. G. (1991). Basic Acoustic Emission, Nondestructive Testing Monographs and tracts Vol. 6.

Smith, L. I., (2002). A Tutorial on Principal Component Analysis: Derivation, Discussion, and Singular Value Decomposition, University of California, San Diego.

Stein, S., and Wysession M. (2003). An introduction to seismology, earthquakes, and earth structure, Blackwell Publishing.

Tan, J. F., Bland, H. C., and Stewart, R.R. (2007). Classification of microseismic events from bitumen production at Cold Lake, Alberta, Thesis, Department of Geosciences, University of Calgary, Alberta.

Tan J.F., Bland H.C., and Stewart R.R. (2007): Passive seismic event classification techniques applied to heavy oil production from Cold Lake, Alberta, CSPG CSEG Convention.

Tan J.F., Stewart R.R., and Wong J. (2009): Classification of microseismic events via principal component analysis of trace statistics, CSEG Conference.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J. (1998): A comparison of selected trigger algorithms for automated global seismic phase and event detection, Bulletin of Seismological Society of America, Vol. 88.

Wong J., Han L., Bancroft J.C, and Steward R.R, (2009): Automatic time-picking of first arrivals on noisy microseismic data, CREWES, University of Calgary.

Wong J., Han L., Bancroft J.C., and Stewart R.R. (2009): Automatic time-picking of first arrivals on noisy microseismic data, CSEG Conference Abstract.

This volume is the property of the University of Oklahoma, but the literary rights of the author are a separate property and must be respected. Passages must not be copied or closely paraphrased without the previous written consent of the author. If the reader obtains any assistance from this volume, he must give proper credit in his own work.

I grant the University of Oklahoma Libraries permission to make a copy of my thesis upon the request of individuals or libraries. This permission is granted with the understanding that a copy will be provided for research purposes only, and that requestors will be informed of these restrictions.

NAME ████████████████████_____

DATE ████████████_____

A library which borrows this thesis for use by its patrons is expected to secure the signature of each user.

This thesis by _____ALVARO A. ORTIZ_____ has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS	DATE