1993

A study of noise reduction techniques used to process multiplex gas chromatographic signals

Maxine R. Shao
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Shao, Maxine R., "A study of noise reduction techniques used to process multiplex gas chromatographic signals" (1993). *Master's Theses*. 713.
DOI: <https://doi.org/10.31979/etd.zqvj-5mrk>
https://scholarworks.sjsu.edu/etd_theses/713

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 1356496

**A study of noise reduction techniques used to process multiplex
gas chromatographic signals**

Shao, Maxine Rose, M.S.

San Jose State University, 1993

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

**A STUDY OF NOISE REDUCTION TECHNIQUES
USED TO PROCESS MULTIPLEX GAS
CHROMATOGRAPHIC SIGNALS**

A Thesis

Presented to

The Faculty of the Department of Physics

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Maxine R. Shao

December, 1993

© 1993

Maxine R. Shao

ALL RIGHTS RESERVED

APPROVED FOR THE DEPARTMENT OF PHYSICS

Dr. J. F. Becker
Physics Department,
San Jose State University

Dr. P. Hamill
Physics Department,
San Jose State University

Dr. J.R. Valentin
Solar System Exploration Branch,
NASA Ames Research Center

APPROVED FOR THE UNIVERSITY

Abstract

A Study Of Noise Reduction Techniques Used To Process Multiplex Gas Chromatographic Signals

by Maxine R. Shao

Gas Chromatography (GC) is an established technique used for the analysis of planetary atmospheres. Multiplex Gas Chromatography (MGC) is a refinement of GC because it increases the signal to noise ratio (S/N) of the chromatographic signal, allowing for the determination of low concentration chemical components. The focus of this study was to test deconvolution and filtering techniques to further increase S/N.

Two different deconvolution algorithms were applied to the MGC output and the algorithm most efficient at increasing S/N was selected. A filtering technique called Clean was applied along with cross-correlation deconvolution to further increase the S/N. The filter mathematically subtracted peaks of relatively large intensities from the MGC output. It was shown that the removal of these peaks increased the S/N of the original deconvolution. The actual increase in S/N was so significant that signals were recovered that were previously buried in the noise.

Acknowledgements

I would like to thank my Thesis advisor, Dr. Jose Valentin, for giving me an interesting research topic and for his continual support throughout the project. Many thanks to my other committee members, Dr. Pat Hamill, and Dr. Joe Becker who not only corrected and edited my work but provided necessary emotional support giving me enough confidence to conclude my research in a timely manner. Last but not least, I would like to thank the one person that was paramount in my actually achieving this goal, my loving husband Bin, without his compassion, understanding, and necessary prodding I never would have begun a Masters degree much less completed this research project.

Table of Contents

1. Introduction.....	1
1.1 Gas Chromatography.....	1
1.2 Limitations of Conventional Gas Chromatography.....	5
1.3 Multiplex Gas Chromatography.....	5
1.4 Research Goals.....	7
2. Theory.....	8
2.1 The Deconvolution of the MGC Output.....	8
2.2 Problems with the Fourier Transform Deconvolution Technique....	11
2.3 Problems with the Cross-correlation Deconvolution Technique.....	13
2.4 The Clean Filtering Technique.....	14
3. Experimental Section.....	16
3.1 The MGC Simulation.....	16
3.2 Deconvolution of the Simulated Data.....	20
3.3 Application of the Filtering Technique.....	25
3.4 Data Collection.....	30
4. Experimental Results.....	33
4.1 Results From the Fourier Transform and Cross-correlation Deconvolution.....	33
4.2 Application of the Clean Technique with Cross-correlation.....	43
4.3 Application of the Clean Technique with Fourier Transforms.....	47
5. Conclusion.....	49
5.1 Summary of Results.....	49
5.2 Suggestions for Future Research.....	50
6. Programs.....	52
6.1 Program 1.....	52
6.2 Program 2.....	67
7. Bibliography.....	73

List of Figures

Figure 1	Diagram of chromatographic separation	4
Figure 3.1	Chromatogram used as the impulse response function	18
Figure 3.2	Simulated MGC output	19
Figure 3.3	Fourier transform deconvolution	23
Figure 3.4	Cross-correlation deconvolution	24
Figure 3.5	Cross-correlation deconvolution with the removal of one peak	28
Figure 3.6	Cross-correlation deconvolution with the removal of two peaks	29
Figure 3.7	Signal-to-noise data table	32
Figure 4.1	Graph comparing Fourier transform deconvolution and cross-correlation deconvolution for 50 injections	34
Figure 4.1a	Data table of the values graphed in Figure 4.1	35
Figure 4.2	Graph comparing Fourier transform deconvolution and cross-correlation deconvolution for 62 injections	36
Figure 4.3	Graph comparing Fourier transform deconvolution and cross-correlation deconvolution for 75 injections	37
Figure 4.4	Graph comparing Fourier transform deconvolution and cross-correlation deconvolution for 100 injections	38
Figure 4.5	Graph comparing Fourier transform deconvolution	

	and cross-correlation deconvolution for 125 injections	39
Figure 4.6	Graph comparing Fourier transform deconvolution and cross-correlation deconvolution for 150 injections	40
Figure 4.7	Graph of the error values for the Fourier transform deconvolution for 75 injections	41
Figure 4.8	Graph of the error values for the cross-correlation deconvolution for 75 injections	42
Figure 4.9	Graph comparing cross-correlation deconvolution and cross-correlation deconvolution with the Clean filtering technique for 50 injections	44
Figure 4.10	Graph comparing cross-correlation deconvolution and cross-correlation deconvolution with the Clean filtering technique for 100 injections	45
Figure 4.11	Graph comparing cross-correlation deconvolution and cross-correlation deconvolution with the Clean filtering technique for 150 injections	46
Figure 4.12	Graph comparing Fourier transform deconvolution and Fourier transform deconvolution with the Clean filtering technique for 50 injections	48

Chapter 1

Introduction

1.1 Gas Chromatography

Gas and Liquid Chromatography are both powerful separation techniques used for the analysis of complex mixtures. The difference between chromatography and other physical and chemical separation techniques is that two immiscible phases are brought into contact where one phase is stationary and the other mobile. The sample mixture, introduced into the mobile phase, goes through a series of interactions with the stationary and mobile phases as it is carried through the system by the mobile phase. Interactions take advantage of differences in the physical or chemical properties of the components in the sample. These differences govern the rate of movement (migration) of the individual components under the influence of a mobile phase as it moves through the column containing the stationary phase. Separated components emerge in the order of the increase in interaction between the sample and the stationary phase. The component with the least interaction elutes first and the component with the most interaction elutes last.

The column is the heart of chromatography since it allows many

types of analyses to be done with a single instrument. This versatility is due to the wide choice of materials for the stationary and mobile phases making it possible to separate molecules that only differ slightly in their physical and chemical properties.

There are many types of chromatographic methods and they are classified according to the types of phases used (1). The mobile phase can be gas or liquid, where the stationary phase can only be a liquid or a solid. For example, when the separation involves a simple partitioning between two immiscible liquid phases, one stationary and one mobile, the process is called liquid-liquid (or partition) chromatography (2). For space exploration type of applications the method of gas chromatography is used (GC) where the mobile phase is gas and the stationary phase is a solid. In the GC experiments that are the focus of this Thesis, the mobile phase is an inert gas, usually helium, and the stationary phase is specifically chosen to separate hydrocarbons.

In the GC experiments studied a sample mixture is modulated into a moving gas stream at the beginning of a column. The column, a narrow metal tube, is packed with porous stationary phase. Due to differential migration the components separate into bands located along the length of the column (Figure 1a). Each component that elutes from the column is detected by a general purpose detector. Plotting this response as a function of time yields a series of nearly symmetric peaks (Figure.1b). The plot of

signal intensity as a function of time is called a chromatogram and is used in both qualitative and quantitative analysis. The retention times of the peaks (*i.e.*, the time at maximum intensity) identify the components. The areas under the peaks, along with a calibration curve, are used to calculate the total concentration of each component.

Figure 1a. Illustrates the chromatographic separation of a two component mixture. Figure 1b is the plot of the detectors response as a function of time. The position of maximum peak height is called the retention time, denoted by t_{R1} for the first component and t_{R2} for the second component.

1.2 Limitations of Conventional Gas Chromatography

Gas chromatography has proven to be a very important technique for analyzing gas mixtures. When applied to the Earth's atmosphere it can determine the permanent components plus the quantity of organic pollutants present. GC has also been used on spacecraft missions to Mars (2) and Venus (3) to determine their atmospheric constituents.

The major limitations that arise when conventional gas chromatography is used to analyze planetary atmospheres are intrinsic to the technique itself. Since some components take a long time to migrate through a column, as a planetary probe descends through an atmosphere only several discrete samples can be taken along its path. These samples will only supply information on the few positions where they were taken, all the information on the intervening positions would be unobtainable. The other fundamental problem is detector sensitivity. If the concentration of the sample is lower than the detection limits of the detector the resulting signal could be lost in the noise created by the system and its environment.

1.3 Multiplex Gas Chromatography

The basic difference between conventional GC and Multiplex GC is that instead of a single sample introduced into the chromatograph multiple samples are modulated into the system. Since there is no waiting for each

sample to elute, information can be obtained more frequently, more information is obtained and the sample volume is increased. The output from this process is digitally stored in a computer. This output is a convolution of the input response function function (the chromatogram) and the modulation signal.

The two methods, studied in this work, to deconvolve multiplex signals are cross-correlation (4), and Fourier transforms (5). The algorithms used for these methods affect the accuracy of the results. Even with the complexity of the signal analysis, MGC still offers benefits over conventional GC, with MGC detection limits of the system will be decreased due to an increase in the input signal. Actually the signal-to-noise ratio is proportional to the square root of samples taken (5). This “multiplex” advantage is the major reason for further study in this area, and has been shown useful in trace analysis determinations (6). A second advantage is the possibility to analyze a continuous stream of gas (as opposed to discrete samples), since there is no need to wait for each sample to elute. This important aspect has been employed to determine methane in ambient air (7). A third advantage is that there is the potential for use in commercial applications where continuous on-line monitoring would show chemical variations.

Although MGC is a powerful technique, it is only as powerful as the data analysis which is applied to the output. These techniques need to be

applied carefully and adjusted so that they fit the experimental system in which they are applied. Problems can also arise with regard to sampling rates, input sequences, and other data acquisition considerations.

1.4 Research Goals

The main goal of this research project was to determine how the signal to noise ratio of multiplex gas chromatographic data can be improved. First, the MGC experiment was computer simulated. This simulation yielded a MGC output. Afterward, two deconvolution tests were performed on the output. The deconvolution which gave the best improvement in detection limits was determined experimentally. Next, the goal was to further increase the S/N by use of a filtering technique on the resolved input response function (chromatogram) found from the deconvolution. Finally it was determined how accurately the deconvolution and filtering technique would yield the intensity values of the chromatogram.

Chapter 2

Theory

2.1 The Deconvolution of the MGC Output

Since the purpose behind MGC is to increase detector sensitivity by increasing sample volume, we must make some assumptions about the way in which this additional sample is acquired. Many samples are taken before the elution time of the previous sample is over. It is important that the system will yield the exact same impulse response function for each sample. Therefore it is assumed that the system will remain stationary during the sampling period. Stationary in this case means that the concentration and number of the components in the sample will not change over time. A further assumption employed when deriving the Fourier deconvolution method is that each injection can be considered to be a delta function. The Multiplexed output of the system is linear since it is a superimposition of impulse response functions. In this context our experiment is viewed as:

Input $\rightarrow$ (Column $\Rightarrow$ Detector $\Rightarrow h(\tau)$) $\rightarrow$ Output

Which can be restated mathematically as:

$$(1) \quad \text{Output} = h(\tau) \otimes \text{Input}$$

where the output is a convolution of the impulse response function, $h(\tau)$, and the input. The input is a random sequence, chosen so that no frequencies that could be present would be lost once a Fourier Transform is applied.

Now taking the Fourier transform of equation 1 (Equation 2), and then applying the appropriate transform property (Equation 3) (8) gives equation 4.

$$(2) \quad \text{FT}(\text{Output}) = \text{FT}(h(\tau) \otimes \text{Input})$$

$$(3) \quad \text{FT}(\text{Output}) = \text{FT}(h(\tau))\text{FT}(\text{Input})$$

$$(4) \quad h(\tau) = \text{FT}^{-1} \left\{ \frac{\text{FT}(\text{Output})}{\text{FT}(\text{Input})} \right\}$$

Equation 4 will be referred to as the Fourier Transform method of deconvolution. Now to derive the second deconvolution method, cross-

correlation, the top and bottom are multiplied by the Fourier transform of the input sequence to get;

$$(5) \quad h(\tau) = \text{FT}^{-1} \left\{ \frac{\text{FT}(\text{Output})\text{FT}(\text{Input})}{\text{FT}(\text{Input})^2} \right\}$$

Now if it is assumed that each injection can be approximated as a delta function, the denominator $\text{FT}(\text{Input})$ goes to 1.

$$(6) \quad \text{FT}\{\text{Input}\} = \int_{-\infty}^{\infty} \delta(x)e^{-i\omega x} dx = e^{-i\omega(0)} = 1$$

So equation 5 becomes;

$$(7) \quad h(\tau) = \text{FT}^{-1} \{ \text{FT}(\text{Output})\text{FT}(\text{Input}) \}$$

Now once again applying the transform property

$$(8) \quad h(\tau) = \text{FT}^{-1} \{ \text{FT}(\text{Output} \otimes \text{Input}) \} = \text{Output} \otimes \text{Input}$$

which by definition is

$$(9) \quad h(\tau) = \sum_{t=1}^N \text{Output}(t)\text{Input}(t-\tau)$$

This result for the crosscorrelation function is not normalized.

Normalizing by the factor $1/N$ gives the crosscorrelation function (equation 10) the second deconvolution method applied in this work.

$$(10) \quad h(\tau) = \frac{1}{N} \sum_{t=1}^N \text{Output}(t)\text{Input}(t-\tau)$$

Both equations studied for deconvolution are mathematically similar but the various algorithms used to compute these equations are not. One of the most important aspects of this work is to investigate how two different algorithms, one for the Fourier transform method, and one for cross-correlation compare in terms of accuracy of deconvolution and computational time.

2.2 Problems with the Fourier Transform Deconvolution Technique

There are problems inherent to these deconvolution techniques. If a new component enters the system during the multiplex experiment it will

affect the output in such a way as to destroy the accuracy of the deconvolved chromatogram. When the new component is added to the system it will produce a new impulse response function. In Chapter 1 it was assumed that the impulse response, $h(\tau)$, would remain constant. This assumption was paramount in the derivation of the Fourier transform and cross-correlation deconvolution techniques. In theory multiplex can only be used only for stationary systems (5). Multiplex GC possibly could be applied to a changing system, but it would be necessary to know where in the multiplex output data the sample had changed. The multiplex output could then be deconvolved in sections. There would be an undetermined amount of error involved in the deconvolutions due to the overlapping range of convolution with the two different impulse response functions. So although we do achieve a multiplex advantage by increasing sample volume we diminish accuracy considerably in an unknown way when that sample changes.

Deconvolution by the Fourier transforms rely on the use of a proper input signal. A proper input signal would not contain zero or near zero amplitudes over the frequency range which the chromatographic system transmits. If a near zero frequency is encountered the denominator of equation 4 becomes very small. When back transformed into the time domain the result is a large spike. When this happens the accuracy of the deconvolved chromatogram is greatly reduced. This problem requires

filtering in the frequency domain to eliminate the near zero amplitudes or careful selection of the input signal.

Although Fourier transforms give the most accurate deconvolution, they are so sensitive to the input sequence that they are not always practical to use. It is best to use a random input sequence so that all possible frequencies will be allowed (9). Unfortunately if one wants to do multiplex GC with a small number of injections the chances of having your deconvolution destroyed by perceived periodicity is very high. If one plans on taking several hundred injections then the periodicity of the input sequence is drastically reduced. This was shown to be true experimentally as accuracy in the Fourier transform deconvolution increased with increased injections.

2.3 Problems with the Cross-correlation Deconvolution Technique

Cross-correlation is a less sensitive technique that is easier to understand and program since it is a simple summation. As in the Fourier transform case, linearity is assumed. Equation 10, the cross-correlation formula contains only multiplication, summation, and averaging. These operations will reduce everything except points that repeat themselves in the same sequence (*i.e.*, the unchanging impulse response function) If anything changes, the cross-correlation operation itself will produce noise.

This noise has been termed “correlation noise” by Annino (4).

Theoretically there is a way to adjust the input signal in such a way as to eliminate correlation noise produced by a sample that varies in concentration (10), but when put into practice correlation noise is still present.

Both the Fourier transform and the cross-correlation techniques require a linear system to perform at their best. To understand the amount of error involved in applying these techniques a computer simulation of the MGC experiment was performed. The situation was ideal since the system was kept linear. This simulation supplied an experimental output that was deconvolved with the two deconvolution techniques to learn what other limitations exist and how they can be treated so that S/N could be maximized.

2.4 The Clean Filtering Technique

Assuming that we have a linear system as previously discussed, a filtering technique was devised to be used along with the deconvolution methods. The idea for implementation of this technique came from personal communications with Dr. J.B. Phillips of Southern Illinois University. The idea is similar to a technique called ‘Clean’ used in Pattern Recognition (11,12) and astronomy (13). Simply stated, by removing the affects of large intensities the S/N ratio will be increased. Each component

effects, in some not well understood way, the final outcome of the deconvolution. This is especially seen in the case of cross-correlation when more than one component is present. The effect of a large component nearby is to affect the accuracy of the deconvolution of smaller components. If the larger components can be mathematically removed the noise they contributed will also be removed. This idea will be tested regarding its usefulness for the analysis of MGC data.

Chapter 3

Experimental

3.1 The Multiplex Gas Chromatography Simulation

A computer simulation of the Multiplex GC experiment was created to analyse and compare the Fourier transform and cross-correlation deconvolution techniques. It was necessary to understand which technique best reduced S/N under equivalent experimental conditions. Afterward, the filtering technique would be applied to both deconvolution methods to determine which would yield the most accurate results.

As a basis for the simulation a single injection chromatogram was generated experimentally using conventional gas chromatography. This single injection chromatogram is the impulse response function which we will later be trying to recover. The peaks shown in Figure 3.1 are ethane, propane, and butane. The concentration of each component in the gas mixture was 1000 parts per million (ppm). Since this sample was of such high concentration it provided a very high S/N signal well beyond the detector's detection limits. The experimental data were used instead of simulated data because it provided a real representation of the peak shapes and their retention times.

IDL (Interactive Data Language, Version 2.0, Research Systems, Inc., Boulder, CO) was used as the programming language for the simulation. This high level language contained the algorithm for the discrete form of the Fourier transform, and the random sequence generator which was used for the input sequence and for the simulated noise that was added to the signal.

To generate the multiplex data, the chromatogram (impulse response function) was convoluted by computer with a random input sequence generated by IDL. This convolution, equation 1, is the MGC output. The number of injections corresponded to the number of elements in the random array. Noise is then added on the baseline, which was set at zero for measuring convenience. This array, called the output, is the simulation of experimental MGC data (Figure 3.2). This simulation yields an 'ideal' output since the sample concentration is virtually fixed throughout the experiment. In real experiments, additional sources of error are introduced. These sources originate from slight changes in the concentration of the components during the experiment.

Figure 3.1. The three peak single injection chromatogram used for the MGC simulation.

Peak 1: Ethane, retention time = 85 seconds.

Peak 2: Propane, retention time = 101 seconds.

Peak 3: Butane, retention time = 151 seconds.

Figure 3.2. Simulated MGC Output. The noise added about the baseline has an amplitude of 1000.

3.2 Deconvolution of the Simulated Data

The convolution MGC output (Figure 3.2) was analyzed using both cross-correlation, equation 10, and the Fourier transform relation, equation 4. Cross-correlation was programmed into the simulation experiment as a simple loop (see Program 1, p.59). The discrete form of the Fourier transform was used to compute equation 4. The algorithm for the forward transform, equation 11, and the back transform, equation 12, was obtained from the mathematical library provided in the IDL software package.

$$(11) \quad F(\omega) = \frac{1}{N} \sum_{t=0}^{N-1} f(t) e^{-i2\pi\omega t/N}$$

$$(12) \quad f(t) = \sum_{\omega=0}^{N-1} F(\omega) e^{i2\pi\omega t/N}$$

Both of the deconvolution methods are sensitive to the input sequence, but Fourier transforms are especially sensitive when the signal is periodic. A problem arises when the input sequence is short, even though the sequence is random, if it is too short the transform perceives it as periodic. This problem is reduced with a longer input sequence and for this reason only input sequences containing more than 50 elements were used in this

comparison. Another problem with the Fourier transform relationship occurs when the input sequence yields a near zero amplitude in the frequency domain, resulting in equation 4 blowing up.

$$(4) \quad h(\tau) = \text{FT}^{-1} \left\{ \frac{\text{FT}(\text{Output})}{\text{FT}(\text{Input})} \right\}$$

Near zero frequencies are likely to occur in a random sequence and it is not noticeable until the transform is taken. To eliminate this problem the transform would need to be filtered in the frequency domain to remove the near zero frequencies before it is back transformed.

When selecting a deconvolution method it is necessary to understand how well the method will perform based on the experimental parameters. These parameters include the number of injections, the amount of noise expected to be present in the system (this can be determined by observing the noise signal of the experimental system without a sample present), the type of modulation sequence used, whether or not there are peaks of low concentration present, and whether or not a filtering technique will be applied to the deconvolution outcome. These variables were varied individually to identify which method worked the best under similar circumstances. It is important to notice that the two techniques studied returned noticeably different impulse response functions with all the

parameters listed above kept constant. Figure 3.3 shows the unfiltered Fourier transform deconvolution. A lower intensity peak at retention time = 85 seconds can be noticed, whereas in the cross-correlation deconvolution (Figure 3.4) this peak is not detected.

Both figures illustrate the differences between the two deconvolutions studied. Both deconvolutions were performed on the same MGC data sequence. These figures should not be used to draw conclusions on the nature of the two deconvolutions behavior in general since depending on the random sequence these figures change in non-straightforward ways. To determine how these different parameters actually effect the deconvolutions results a statistical average of many deconvolutions with fixed parameters was performed. The two parameters, number of injections (the number of elements in the random sequence) and the noise intensity, were varied and their effect on the deconvolutions was studied.

Figure 3.3. The Fourier transform deconvolution of the simulated MGC output. 50 injections were used to form the original output signal. The original output signal (Figure 3.2) had a noise intensity of 1000.

Figure 3.4. The cross-correlation deconvolution. Same parameters as in the Fourier transform deconvolution shown in Figure 3.3. The small peak at 85 seconds is not resolved in this case.

The maximum relative intensity of the three peaks used in the simulation are: Peak 1 at 85 seconds, $I=416.45$, Peak 2 at 101 seconds, $I=2404.97$, Peak 3 at 151 seconds, $I=2322.47$. These values of maximum intensity are from the initial data file used to create the simulation (Figure 3.1). Intensity values are given in arbitrary units since in a real MGC experiment a calibration curve for each component would need to be created using the experimental system. Using the calibration curve these intensities would then be able to be converted into their actual concentration in parts per million. These values were used as the basis to determine how efficiently the deconvolution techniques perform. From now on they will be referred to as the 'known' values, or the given values. As one can see from Figures 3.3 and 3.4 the Fourier transform deconvolution resolved these three peaks more accurately.

3.3 Application of the Filtering Technique

The filtering technique to be applied after the deconvolution procedure had its origins in astronomy. In that field it is called 'Clean' so this term will be adopted for reference when discussing it. Although this filter has never been applied to GC signals, it was simple and straightforward enough to be applied to MGC data. The premise is to remove the effect of neighboring peaks to reduce their influence on the smaller components. Once these larger peaks are removed we are basically

looking for a reduction in noise. To apply this technique it is necessary to mathematically eliminate the peaks with large intensity in order of the largest first and look for increased S/N.

To do this the peak to be removed is first isolated on the deconvolved chromatogram. This is done by identifying the zero points on either side of the peak in the initial deconvolution, then this range of points is selected from the initial impulse response function used for the MGC simulation. Finally this peak shape is copied, then multiplexed with the same random sequence, giving a MGC output of this one peak. This single peak MGC output was then subtracted, point by point from the original MGC output, thereby eliminating this peak from the data. This new MGC output when deconvolved will have the peak successfully removed along with its effects on the other peaks in the chromatogram.

There are problems in this method if it is not applied carefully. Sometimes the deconvolved peak shape itself contains noise. If this noisy peak shape is used to construct a single peak MGC output there is a poor chance for noise reduction after the final deconvolution. For this reason the noisy peak shape was not used to create the single peak MGC output. In a real experiment it would be helpful if not necessary to minimize the difference between the impulse response function and the shape of the function given by the deconvolution. The first deconvolution (Figures 3.3 and 3.4) was used only to determine at which points the peak shapes should

be cut off. Figure 3.5 shows the deconvolution (in this case cross-correlation was used) after the largest peak at $t = 101$ seconds) had been removed by this process. Comparing this to Figure 3.4 it is seen that a small peak at 85 seconds can start to be distinguished against the noise. Figure 3.6 shows the result after both large peaks were removed. Here it can be seen that the S/N is significantly improved and in fact the intensity of the small peak was determined, in this case to be 413.45, only one percent from the known value of 416.45. Although this looks impressive it will be shown in Chapter 4 that this technique can work consistently.

In comparing Figures 3.4 and 3.6 it is obvious that the peaks at $t = 101$ and $t = 151$ seconds had not been entirely removed and can be seen as remnants appearing in their former positions. When not counting these two peaks as noise the signal to noise ratio for Figure 3.6 would be roughly equal to 3 (small peak intensity/maximum noise intensity) whereas previously the small peak at $t = 85$ seconds was completely buried in the noise.

Figure 3.5. Using cross-correlation deconvolution after removal of the largest peak, at $t = 101$ seconds. A small reduction in noise can be seen when comparing to Figure 3.4.

Figure 3.6. Cross-correlation deconvolution after removal of the two large peaks at $t=101$ and $t=151$ seconds. Here the noise is significantly reduced as compared to Figure 3.4. It is also evident that these two peaks were not completely removed.

3.4 Data Collection

All data were collected directly from the simulation programs. Each simulation was made then the output was deconvolved and filtered. The data were recorded in tables then entered into the Macintosh program, Statview which computed the mean values and the standard deviations. A statistical approach was used to evaluate which of the two deconvolutions produced the most accurate results (*i.e.*, which deconvolved the impulse response function with the least amount of error). Ten trial simulations were done with fixed noise amplitude and number of injections. The resulting MGC output was then deconvolved using both the Fourier transform and cross-correlation deconvolution methods. The one variable that changed for each simulation was the pseudo random injection sequence which was generated in IDL.

The effect of increasing the number of injections was studied by holding the noise constant and repeating the simulation with six different injection amounts ranging from 50 to 150 injections. It took three to six hours to complete one simulation using 150 injections, along with its deconvolutions, depending on how many users were logged onto the MicroVax system. Due to this length of time it was decided to limit the number of injections to 150. The relative noise intensity ranged from 140 to 2000. Recall that the peak intensities used in the simulation were 415,

2405 and 2322. With the large noise values, from 1500 to 2000 the small peak rarely was resolved. The filtering technique Clean was applied in conjunction with cross-correlation in all cases to determine its effectiveness.

The signal to noise was computed separately for each peak, for example if the noise intensity is 2000 and we consider each of the three peaks, the one with height 416 would have $S/N = 416/2000 = 0.21$, the peak with height 2405 would have $S/N = 2405/2000 = 1.2$, and the peak with height 2322 would have $S/N = 2322/2000 = 1.16$. In this way 15 discrete signal to noise ratios were found (Figure 3.7).

		Noise Intensity				
		139	200	313	600	2000
Peak Intensity	Peak #1 (416.45)	3.00	2.08	1.33	0.69	0.21
	Peak #2 (2404.97)	17.3	12.0	7.68	4.01	1.20
	Peak #3 (2322.47)	16.7	11.6	7.42	3.87	1.16

Figure 3.7. Signal-to-noise values for the three peaks as a function of the noise added to the simulated data. The peak intensity refers to the maximum peak intensity. These are the known values determined from the function used to create the simulation (Figure 3.1).

All results will be presented in terms of the percent difference between the known value and the corresponding value found from the deconvolution or the combined deconvolution and filtering technique. Each percent difference is taken with respect to the mean value found from the ten runs using the fixed parameters of noise and number of injections. The standard deviation about the mean will be presented as error bars on the graphical display of the data. For this study all of these computations were performed for each injection quantity.

Chapter 4

Results

4.1 Results from the Fourier Transform and Cross-correlation Deconvolution.

This initial set of graphs compares how accurately the two deconvolution methods, Fourier transforms and cross-correlation, computed the impulse response function. The first set of graphs, Figures 4.1 through 4.6, show how the error changes with injection amount. The error in resolution is plotted against the signal to noise ratio. In the region between S/N of 0 and 4 the Fourier transform deconvolution is more accurate than the cross-correlation deconvolution by 2% to as much as 30% in some cases.

Figures 4.7 and 4.8 include the standard deviation on the data set of 75 injections. This is typical of the deviations in these errors. It would be useful to run a larger data set (30 trials for each set state of parameters) to reduce these deviations.

Figure 4.1. Percent error as a function of signal to noise for 50 injections. This graph depicts the percent error between the actual peak intensity and the intensity found by each deconvolution method. Each data point represents the mean value of the error taken from 10 different computational trials. The random sequence was the only variable that differed on each trial.

S/N	FFT	SD	Xcor	SD	Clean	SD
.208	90.568	29.827	78.000	36.578	48.170	36.409
.694	49.220	31.900	39.086	33.704	8.149	9.884
1.162	13.734	11.174	9.388	9.434	11.092	8.096
1.202	13.004	13.306	8.256	7.605		
1.340	9.246	6.621	21.616	22.164	10.755	15.842
2.080	10.379	7.633	24.576	23.019	5.297	3.928
3.000	5.662	4.029	33.684	29.769	4.110	3.425
3.888	3.668	3.668	8.996	4.923	10.765	2.605
4.000	3.408	3.461	5.502	4.416		
7.440	2.336	2.055	11.320	3.225	11.038	1.480
7.700	3.169	1.924	7.794	4.693		
11.620	1.562	1.205	12.924	2.664	12.230	1.083
12.020	1.358	1.131	8.664	3.163		
16.700	.744	.576	9.955	4.025	10.448	1.653
17.300	1.153	.692	6.175	3.653		

Figure 4.1a. Data table used to generate Figure 4.1. The first column, S/N, are the data values for the x-axis. The values in columns FFT, Xcor, and Clean represent the mean value of the error found when a peak was recovered at that particular S/N for the Fourier transform deconvolution, the cross-correlation deconvolution, and the combined cross-correlation deconvolution and Clean filtering technique. The columns entitled SD give the standard deviation for each column to their immediate left.

Figure 4.2 Percent error as a function of signal to noise for 62 injections. Same caption as in Figure 4.1. The second figure in a series of six.

Figure 4.3 Percent error as a function of signal to noise for 75 Injections. Same caption as in Figure 4.1. The third figure in a series of six.

Figure 4.4. Percent error as a function of signal to noise for 100 injections. Same caption as in Figure 4.1. The fourth figure in a series of six.

Figure 4.5. Percent error as a function of signal to noise for 125 injections. Same caption as in Figure 4.1. Fifth figure in a series of six.

Figure 4.6. Percent error as a function of signal to noise for 150 injections. Same caption as in Figure 4.1. Last figure in a series of six 150 injections was the maximum number of injections used in this simulation.

Figure 4.7. Percent error as a function of time for the Fourier transform deconvolution with 75 injections. Error bars represent the standard deviation.

Figure 4.8. Percent error as a function of time using cross-correlation deconvolution with 75 injections. The error bars represent the standard deviation. Comparing to Figure 4.7 for small signal to noise ratios the Fourier transform deconvolution method yields a significantly smaller error.

4.2 Application of the Clean Technique with Cross-correlation.

When the study of these deconvolution techniques began, this filtering technique was only applied to cross-correlation. It was very effective in significantly reducing the noise intensity on the cross-correlation deconvolution. A vast quantity of data was collected to prove the usefulness of this technique when applied with cross-correlation, but it was not tested with the Fourier transform deconvolution method. Preliminary tests were performed to evaluate the effectiveness of the filtering technique used with Fourier transforms, but the results were a disappointment and the study was not continued. Section 4.4 contains discussion of the possible reasons for this failure and provide suggestions as to how this technique might be applied in future work.

This section will display the usefulness of the 'Clean' technique on the cross-correlation deconvolution. The three Figures 4.9, 4.10, and 4.11 show the dramatic reduction in error when using this filtering technique compared to using only cross-correlation.

Figure 4.9. Deconvolution using the cross-correlation alone compared with the cross-correlation and the filtering technique. 50 injections were used. In the range below $S/N = 4$ the application of the filtering technique increased the accuracy of peak detection.

Figure 4.10. Same caption as in Figure 4.9. 100 injections were used.

Figure 4.11. Same caption as in Figure 4.9. 150 injections were used.

4.3 Application of the Clean Technique with Fourier Transforms

The application of the Clean filtering technique was applied to the Fourier transform deconvolution, but was not effective. When the filtering technique was applied there was no reduction in noise. Furthermore the transform created unforeseen problems when the peaks were subtracted. When looking at the results of the deconvolution after a peak was removed sometimes the noise shape was completely different. At times the filtering technique caused a loss in accuracy. In general no increase in detection limits was found. Instead the noise amplitude remained constant as the peaks are removed. Figure 4.12 compares the filtered Fourier transforms with the Fourier transforms alone. This data set was collected in the same manner as the data collected with Cross-correlation. Program 2 was used to create the MGC simulation, give the MGC output, and perform the deconvolution of the MGC output. Since Fourier transforms outperformed cross-correlation when applying the first deconvolution, it was assumed that when they were applied along with the filtering technique that detection limits would have been increased. This area warrants further investigation to determine why these techniques did not work in conjunction.

Figure 4.12. Percent error as a function of signal to noise for 50 injections. Comparison of the Fourier transform technique with the filtered Fourier transforms. At the point where $S/N = 3$ the filtering technique actually destroys the original accuracy at which the peaks were determined.

Chapter 5

Conclusion

5.1 Summary of Results

The main goal of this work was to apply various mathematical techniques to Multiplex Gas Chromatographic data to increase the signal to noise ratio. This study determined which of the two deconvolution methods (Fourier transform or cross-correlation), increased S/N most significantly. When S/N was larger than 2 the Fourier transform deconvolution was the preferred method, yielding consistently lower recovery errors than cross-correlation. S/N is typically greater than 2 unless analyzing atmospheric components. In the range where S/N is less than 2, cross-correlation in conjunction with the Clean filtering technique out performed both deconvolution techniques. Clean can recover peaks with S/N of 0.2 with only a 20 percent error. If one has no expectation or interest in components of concentration lower than 2, then the Fourier method alone should be used to resolve the impulse response function. On the other hand if one is doing a MGC experiment to find low concentration components which could be buried in the system noise, then cross-correlation and the Clean technique should be used.

5.2 Suggestions For Future Research

A future improvement to the Fourier transform deconvolution technique would be to increase the transforms accuracy in the small S/N range. This is worthy of investigation since Fourier transforms work so effectively in the higher range of S/N. Either a filtering technique or an adjustment of the transforms could potentially decrease recovery error. Possibly a special input signal could also be chosen that would greatly increase the S/N. It is suspected that the Clean filtering technique failed when applied with Fourier transform deconvolution since Clean could only eliminate a particular type of noise present only in the cross-correlation deconvolution. A future test to determine whether Clean would work when applied with Fourier transforms could make use of a matched filter to isolate and accurately replicate the deconvolved peak shape before subtraction.

The Clean technique did work extremely well with cross-correlation deconvolution. An added improvement would be to remove the peaks in a systematic way making sure to completely eliminate each peak. Also, since one would be using the Clean technique to recover small peaks, all the large peaks should be removed at once to save time.

The problem of changing chemical composition was never discussed in this work since we assumed that the impulse response function (chromatogram) would stay constant. Both deconvolution methods were

derived under this assumption. This research needs to be extended to be able to include changing sample composition. In real atmospheric studies the sample could change in both composition and concentration. These deconvolution methods could presently be used for qualitative applications where a changing sample is involved. As an example, multiplex could be used in a chemical plant where one could be continuously multiplexing air samples. The deconvolution of sections of the output would show the presence of any new components. This idea could be incredibly important as an early warning system in a number of different manufacturing situations. Therefore, multiplex gas chromatography is not only useful for extraterrestrial research, it also has the potential to be a lifesaving monitoring technique.

6.1 Description of Program 1

Program 1 simulates the MGC experimental data then applies both deconvolution techniques to the data. After the deconvolutions have been analyzed, Program 1 then applies the Clean technique along with the cross-correlation deconvolution. This interactive program allows the user to determine the deconvolved peak intensities and then use those values in the application of the Clean technique. In the application of the Clean technique an unlimited number of peaks can be removed. Error estimates are not performed by this program. All errors must be calculated by the user from the values determined by the programs computations.

```

;          FAST.PRO
;
;          04/1991
;
;          * Modified by Maxine Shao 08/91
;
;  Description:
;  An IDL program to read in CHROMATOGRAM of one injection
;  MGC experiment, then simulate the case of N RANDOM
;  INJECTIONS
;  EXPERIMENT with low concentrations and noise added; then
;  try to recover the chromatogram using cross-correlation.
;  Then do template matching on that data over and over until
;  you have recovered all the peaks.
;
; *****
; SET_PLOT,'TEK' ; Set plotting characteristics of the terminal
; *****
;  DATA AREA
; *****
; FILENAME= ' '
; CSTORE= ' '
; TSTORE= ' '
; CHEC=' '
; PRE=' '
; MGCFILE=' '
; CANOTHER=' '
; CINTENS=' '
; FINTENS=' '
; FANOTHER=' '
; PLOOK=' '
; PINTENS=' '
; PANOTHER=' '
; PINTENS=' '
; TANOTHER=' '
; TINTENS=' '
; ELIM=' '
; INDEXX=0
; AHA=0

```

```

;
;
;
;
;*****
;      MAIN PROGRAM
;*****
; Read in the filename of the initial file.
;
;FILENAME='6MAR3R'
;READ,'Enter Original Data Filename ',FILENAME
;
; Read in the number of data points in the initial file. Define the
;   array RAW to be a floating point array which is SIZE points long.
;
;SIZE=900
;READ,'Enter Number Of Data Points In Files ',SIZE
RAW=FLTARR(SIZE)+0.0
;
; This outputs user MMI, says to close unit 1 if there is an error,
; opens unit 1 while looking for the output data file,
; reads the file into the array RAW, and closes unit 1.
;
PRINT,'Reading(',FILENAME+'.DAT',')'
ON_IOERROR,LABEL1
OPENR,1,FILENAME+'.DAT' &READF,1,RAW
LABEL1: &CLOSE,1
;
; Plot the array RAW which has the initial data in it. Print command
;   to the baseline, and read in the starting point from which to
;   calculate the zero.
;
;PLOT,RAW
;PRINT,'You Need To ZERO The Baseline '
M=300
;READ,' To Do This Enter Starting Point To Calculate Baseline (Integer)
',M
;
; This is to zero the baseline of chromatogram. It takes the total of all
;   the intensities starting at the initial point previously defined

```

```

;    and divides it by the number of intensities considered. Then it
;    subtracts this average from the initial data. This is done twice
;    because the second time fine tunes the adjustment. Then the re-
;    adjusted initial data is plotted.
;
AVERAGE=(TOTAL(RAW(M:(SIZE-1)))/(SIZE-1-M))
RAW=RAW-AVERAGE
AVERAGE=(TOTAL(RAW(M:(SIZE-1)))/(SIZE-1-M))
RAW=RAW-AVERAGE
;PLOT,RAW
;PRINT,'This Is The Chromatogram After ZEROING The Baseline '
;
; Again is a label (or marker) which the program uses to return to
;    if the user wants to do run the program again but using the
;    same original data when the program is done.
;
AGAIN:
;
PRINT,'***** Generating Multiplex Data *****'
;
; This asks for the user to input the number of injections for the
;    purpose of creating a multiplexed output. If the user were
;    using experimentally multiplexed data, this step would not
;    be needed. NU is set to NUU-1 for the purpose of setting up
;    do loop structures later in the program, since a loop usually
;    goes from 0 to some number.
;
READ,'Enter The NUMBER Of Injections ',NUU
NU=NUU-1
;
; Define Variables.
;
T=FLTARR(NU)+0.0
DELAY =FLTARR(NU)+0.0
SPACING=INTARR(NU)
;
; Read in the range of spacing between injections which is allowed.
;    In chromatography, the minimum is generally equal to the
;    width-at-half-maximum of the widest peak. This would also

```

```

;    not be needed if the data were experimentally multiplexed.
;
MA=50
MI=10
;READ,'Enter The MAXIMUM Spacing Between Injections (seconds)
',MA
;READ,'Enter The MINIMUM Spacing Between Injections (seconds) ',MI
;
; Calculate the spacing based on the MINimum and the MAXimum spacing
;    values allowed using a RANDOM number generator supplied by
;    IDL. Then load them into a Time array.
;
SPACING=(FIX((MA-MI)*RANDOMU(S,NU)))+FIX(MI)
FOR H=1,NU DO BEGIN
    HH=H-1
    T(HH)=SPACING(HH)
ENDFOR
;
; Read in the unit of time which each data point represents, the
;    maximum amount of noise to the data, and the maximum length
;    of the chromatogram.
;
TUNIT=1
;READ,'Enter Time Units Of Chromatogram (seconds) ',TUNIT
READ,'Enter The MAXIMUM Noise To Be Added ',MAXI
UMBALA=M
;READ,'Enter The MAXIMUM Guessed Length Of Chromatogram
',UMBALA
;
; This is the calculates the number of delay channels before each
;    subsequent injection in terms of the data rate (TUNIT) as
;    well as the total time of delay.
;
DELAY=T/TUNIT
TODELAY=TOTAL(DELAY)
;
; Define Variables. EE will be the array with the cross-correlated
;    chromatogram, and then it will become the array with the
;    template-matched chromatogram. TEMPLATE is the template which

```

```

;      is used to check through the multiplexed data for the peaks.
;
;
;
INDEX=SIZE+TODELAY
EE=FLTARR(INDEX)
TEMPLATE=FLTARR(INDEX)
AA=FLTARR(INDEX)
SPAC=INTARR(INDEX)
;
; Read in the peak of interest from the initial data to the template
;   file.
;
FOR I=0,SIZE-1 DO TEMPLATE(I)=RAW(I)
;
; This is the array of subsequent chromatogram.
;
FOR J=0,NU-1 DO BEGIN
    BOB=AHA+DELAY(J)
    FOR I=0,SIZE-1 DO
        TEMPLATE(I+BOB)=TEMPLATE(I+BOB)+RAW(I)
    AHA=AHA+DELAY(J)
ENDFOR
;
; This is to add noise to the template about the zero to make it more real.
;
;
SPAC=(FIX((MAXI)*RANDOMU(S,INDEX)))
SPAC=SPAC-(MAX(SPAC)-MIN(SPAC))/2.
TEMPLATE=TEMPLATE+SPAC
PLOT,TEMPLATE
PRINT,'This Is The Template With Noise Added '
;
AA(0)=1
FOR I=0,NU-1 DO BEGIN
    INDEXX=INDEXX+SPACING(I)
    AA(INDEXX)=1
ENDFOR

```

```

;
PRINT,'***** Starting Fourier Transform Program *****'
;
;
;   THIS IS THE MATHEMATICALLY CORRECT DISCRETE
FOURIER TRANSFORM
;
DATA=FFT(TEMPLATE,-1)
TEMPLATE=FFT(DATA,1)
FAA=FFT(AA,-1)
FINAL=FFT(DATA/FAA,1)/(INDEX)
PLOT,FINAL
;
; This plots the final data over a user specified range and then reads the
;   user specified intensities.
;
FLOOK:
READ,'Enter Range Over Which To Look For Peak #,# (Integer)
',FLOOK1,FLOOK2
PLOT,FINAL(FLOOK1:FLOOK2)
READ,'Do You Want An Intensity Reading? (Y/N) ',FINTENS
IF (FINTENS EQ 'Y') THEN BEGIN
    READ,'Enter Point Of Peak Intensity (Intgr) ',FPOINT
    PRINT,' Intensity at ',FPOINT,' = ',FINAL(FPOINT)
ENDIF
READ,'Do You Want Another Look ? (Y/N) ',FANOTHER
IF (FANOTHER EQ 'Y') THEN GOTO,FLOOK
;
PRINT,'***** Starting Cross-Correlation Program *****'
;
; This is to set up the data array for use with cross-correlation.
;
TEMPLATE=FLOAT(TEMPLATE)
MTEMPLATE=TEMPLATE
CTEMPLATE=FLTARR(2*INDEX)
CTEMPLATE(0:INDEX-1)=TEMPLATE
;
; This is the UNBIASED form of the CROSS-CORRELATION summation
;   assuming NON-PERIODIC BOUNDARY CONDITIONS. We use

```

NON-

```
; PERIODIC BOUNDARY CONDITIONS because we assume that the
; sample could be changing with time and that the last injection
; could come at the last sample point, not allowing enough
; time to measure the final peaks coming out of the column.
; We use the UNBIASED form of the cross-correlation because
; as the cross-correlation progresses, it multiplies say
; 20 numbers in the array together and totals them, and then
; 19, and then 18, and 17,....,and 4.... So you cannot always
; divide by the total number of data points in the array.
```

```
;
FOR J=0,INDEX-1 DO EE(J)=TOTAL(AA*CTEMPLATE(J:INDEX-
1+J))/(INDEX-J)
```

```
;
; The reason for this scaling factor is that the AA array is a series
; of randomly spaced delta functions. Each time you get a 1 in
; AA(K), EE(J)=TEMPLATE(K). This scaling must be performed
AFTER
```

```
; the cross-correlation summation in order to maintain the
; mathematical integrity of the cross-correlation function.
```

```
;
EE=EE*INDEX/NUU
```

```
;
; This is to correct the baseline moving due to cross-correlation method
; using an average of the last 20 points in the chromatogram.
; Then the program plots the final chromatogram calculated using
; the cross-correlation method.
```

```
;
CAVERAGE=(TOTAL(EE(UMBALA-21:(UMBALA-1)))/(UMBALA-1-
(UMBALA-21)))
```

```
EE=EE-CAVERAGE
```

```
CAVERAGE=(TOTAL(EE(UMBALA-21:(UMBALA-1)))/(UMBALA-1-
(UMBALA-21)))
```

```
EE=EE-CAVERAGE
```

```
;
PLOT,EE
```

```
PRINT,'This Is Chromatogram From The Cross-Correlation Method '
```

```
;
; This plots the final data over a user specified range and then reads the
```



```

;    user specified intensities.
;
CLOOK:
READ,'Enter Range Over Which To Look For Peak #,# (Integer)
',CLOOK1,CLOOK2
PLOT,EE(CLOOK1:CLOOK2)

READ,'Do You Want An Intensity Reading? (Y/N) ',CINTENS
IF (CINTENS EQ 'Y') THEN BEGIN
    READ,'Enter Point Of Peak Intensity (Intgr) ',CPOINT
    PRINT,' Intensity at ',CPOINT,' = ',EE(CPOINT)
ENDIF
READ,'Do You Want Another Look ? (Y/N) ',CANOTHER
IF (CANOTHER EQ 'Y') THEN GOTO,CLOOK
;
;*****
; Now the program begins with the TEMPLATE MATCHING.
;*****
;
PRINT,'***** Starting Template Matching Program *****'
;
AHA=0
;
CHECK:
TAGAIN:
;
; Define variables.
;
PEAK=FLTARR(SIZE)+0.0
PEAK1=FLTARR(SIZE)+0.0
EE=FLTARR(INDEX)+0.0
TEMPLATE1=FLTARR(INDEX)+0.0
TEMPLATE2=FLTARR(INDEX)+0.0
;
READ,'Enter The XCORR Peak Height (Real) ',X
READ,'Enter The STANDARD Peak Height (Real) ',Q
;
READ,'Enter The BEGINNING,END Of Peak To Be Deleted (Integer)

```

```

',END1,END2
READ,'Enter The Retention Time Of Peak To Be Deleted (Integer)
',TARA
PRINT,' '
;
; Asks you if the questions you answered were correct, if not then
; lets you re-answer ALL of them.
;
READ,'Is This All Correct ? (Y/N) ',CHEC
IF (CHEC EQ 'N') THEN GOTO,CHECK
;
; Calculate ratio of peak height from cross-correlation to that
; from the actual data.
;
TATA=X/Q
;
; This is to create a template containing only one of the peaks.
;
PEAK(END1:END2)=RAW(END1:END2)
;
; This is to make sure that every peak that shall be used for
; "one peak only" templates will have the proper retention times.
;
PEAK=TATA*PEAK
HAHA=SORT(PEAK)
HABIG=HAHA(SIZE-1)
FOR J=0,20 DO PEAK1(TARA-5+J)=PEAK(HABIG-5+J)
;
; This is the array of subsequent multiplexed output.
;
AHA=0
FOR J=0,NU-1 DO BEGIN
    BOB=AHA+DELAY(J)
    FOR KI=0,SIZE-1 DO
        TEMPLATE1(KI+BOB)=TEMPLATE1(KI+BOB)+PEAK1(KI)
    AHA=AHA+DELAY(J)
ENDFOR
;
; The TEMPLATE2 equals the multiplexed output minus the peak we are

```

```

;      trying to get rid of.
;
;
TEMPLATE2=TEMPLATE-TEMPLATE1
;
; This is to set up the data array for use with cross-correlation.
;
TEMP=TEMPLATE2
TEMPLATE2=FLTARR(2*INDEX)
TEMPLATE2(0:INDEX-1)=TEMP
;
; This is the UNBIASED form of the CROSS-CORRELATION summation
;      assuming NON-PERIODIC BOUNDARY CONDITIONS. We use
NON-
;      PERIODIC BOUNDARY CONDITIONS because we assume that the
;      sample could be changing with time and that the last injection
;      could come at the last sample point, not allowing enough
;      time to measure the final peaks coming out of the column.
;      We use the UNBIASED form of the cross-correlation because
;      as the cross-correlation progresses, it multiplies say
;      20 numbers in the array together and totals them, and then
;      19, and then 18, and 17,....,and 4.... So you cannot always
;      divide by the total number of data points in the array.
;
FOR J=0,INDEX-1 DO EE(J)=TOTAL(AA*TEMPLATE2(J:INDEX-
1+J))/(INDEX-J)
;
; The reason for this scaling factor is that the AA array is a series
;      of randomly spaced delta functions. Each time you get a 1 in
;      AA(K), EE(J)=TEMPLATE(K). This scaling must be performed
AFTER
;      the cross-correlation summation in order to maintain the
;      mathematical integrity of the cross-correlation function.
;
EE=EE*INDEX/NUU
;
; This is to correct the baseline moving due to cross-correlation method
;      using an average of the last 20 points in the chromatogram.
;      Then the program plots the final chromatogram calculated using
;      the cross-correlation method.

```

```

;
TAVERAGE=(TOTAL(EЕ(UMBALA-21:(UMBALA-1)))/(UMBALA-1-
(UMBALA-21)))
ЕЕ=ЕЕ-TAVERAGE
TAVERAGE=(TOTAL(EЕ(UMBALA-21:(UMBALA-1)))/(UMBALA-1-
(UMBALA-21)))
ЕЕ=ЕЕ-TAVERAGE
;
PLOT,ЕЕ

;
; This plots the final data over a user specified range and then reads the
;   user specified intensities.
;
    ALOOK:
    READ,'Enter Range To Look For Peak #,# (Integer)
',LOOK1,LOOK2
    PLOT,ЕЕ(LOOK1:LOOK2)
    READ,'Do You Want An Intensity Reading? (Y/N) ',TINTENS
    IF (TINTENS EQ 'Y') THEN BEGIN
        READ,'Enter Point Of Peak Intensity (I) ',POINT
        PRINT,' Intensity at ',POINT,' = ',ЕЕ(POINT)
    ENDIF
    READ,'Do You Want Another Look ? (Y/N) ',TANOTHER
    IF (TANOTHER EQ 'Y') THEN GOTO,ALOOK

;
;
TEMPLATE=TEMPLATE2
; Read in whether or not the user wants to continue and eliminate
;   another peak.
;
READ,'Do You Want To Eliminate Another Peak ? (Y/N) ',ELIM
IF (ELIM EQ 'Y') THEN GOTO,TAGAIN
;
;*****
;
;   MATRIX.PRO
;*****
;

```

```

PRINT,'***** Starting Matrix Program *****'
;
RENEW: MSTORE= ' '
;
; Read in the number of peaks.
;
READ,'Enter Number Of Peaks (Integer) ',NUM
;
; Define variables.
;
MPEAK=INTARR(NUM*2+1)
EA=DBLARR(NUM+1,SIZE)+0.0
F=DBLARR(NUM)+0.0
RT=DBLARR(NUM)+0.0
;
; Read in the retention times and the start and end points of each of
;   the peaks.
;
FOR I=1,NUM DO BEGIN
    READ,'From Cross-Correlation, Enter The Retention Time ',TARA
    READ,'Enter Start, End of Current Peak ',P1,P2
    MPEAK(2*I-1)=P1 &MPEAK(2*I)=P2
    RT(I-1)=TARA
ENDFOR
;
TEMPLA1=DBLARR(INDEX)+0.0
TEMPLA2=DBLARR(INDEX)+0.0
TEMPLA3=DBLARR(INDEX)+0.0
KKK=DBLARR(NUM,INDEX)+0.0
;
; This reads in each portion of the original data which contains a peak
;   and reads it into the N-dimensional matrix EA (where N is the
;   number of peaks).
;
FOR I=1,NUM DO EA(I,MPEAK(2*I-1):MPEAK(2*I))=RAW(MPEAK(2*I-1):MPEAK(2*I))
;
; This begins the loop which makes the template.
;

```

```

FOR I=1,NUM DO BEGIN
  HAHA=DBLARR(SIZE)
  HAHA=SORT(EA(I,*))
  HABIG=HAHA(SIZE-1)
  F(I-1)=RT(I-1)
  RAW21=DBLARR(SIZE)+0.0
  FOR J=0,15 DO RAW21(F(I-1)-5+J)=EA(I,HABIG-5+J)
  TEMPLATE1=DBLARR(INDEX)+0.0
  FOR KI=0,SIZE-1 DO TEMPLATE1(KI)=RAW21(KI)
  AHA=0
  FOR J=0,NU-1 DO BEGIN
    BOB=AHA+DELAY(J)
    FOR KI=0,SIZE-1 DO
      TEMPLATE1(KI+BOB)=TEMPLATE1(KI+BOB)+RAW21(KI)
    AHA=AHA+DELAY(J)
  ENDFOR
  IF (I EQ 1) THEN TEMPLA1=TEMPLATE1
  IF (I EQ 2) THEN TEMPLA2=TEMPLATE1
  IF (I EQ 3) THEN TEMPLA3=TEMPLATE1
ENDFOR
;PLOT, TEMPLATE1
;
; This reads in all the peak templates in the the array KKK.
;
;
FOR I=0,NUM-1 DO BEGIN
  FOR J=0,INDEX-1 DO BEGIN
    IF (I EQ 0) THEN KKK(I,J)=TEMPLA1(J)
    IF (I EQ 1) THEN KKK(I,J)=TEMPLA2(J)
    IF (I EQ 2) THEN KKK(I,J)=TEMPLA3(J)
  ENDFOR
ENDFOR
;
; Do the template matching.
;
KIK=DBLARR(INDEX,NUM)+0.0
KIK=TRANPOSE(KKK)
OP=KKK#KIK
INVOP=INVERT(OP)

```

```

MK=DBLARR(1,NUM)
MK=MTEMPLATE#KIK
XXX=DBLARR(NUM)+0.0
XXX=MK#INVOP
;
; This is to reconstruct the chromatogram from the matrix XXX.
;
RECONSTRUCT=DBLARR(SIZE)+0.0
FOR I=1,NUM DO BEGIN
  FOR J=0,16 DO RECONSTRUCT(F(I-1)-5+J)=EA(I,RT(I-1)-
5+J)*XXX(I-1)
ENDFOR
;
; This plots the reconstructed chromatogram.
;
PLOT,RECONSTRUCT
PRINT,'this is the CONSTRUCTED chromatogram from the found peaks
heights'
;
PRINT,'PEAK AT 85 = ',RECONSTRUCT(85)
PRINT,'PEAK AT 101 = ',RECONSTRUCT(101)
PRINT,'PEAK AT 151 = ',RECONSTRUCT(151)
;
; Reads in if the user wants to continue with another data set.
;
; If CONTINUE equals 'Y', then start program over.
CONTINUE=' '
READ,' Do You Want To Run The Program Again ? (Y/N) ',CONTINUE
IF (CONTINUE EQ 'Y') THEN GOTO,AGAIN

```

6.2 Description of Program 2

Program 2 simulates the MGC experimental data then applies the Fourier transform deconvolution technique to the data. After the deconvolution has been analyzed, Program 2 then applies the Clean technique along with the Fourier transform deconvolution. This interactive program allows the user to determine the deconvolved peak intensities and then use those values in the application of the Clean technique. In the application of the Clean technique an unlimited number of peaks can be removed. Error estimates are not performed by this program. All errors must be calculated by the user from the values determined by the programs computations.


```

;          MAX.Pro
;
;          M.R.Shao  7/6/93
;
;
SET_PLOT,'TEK' ; Sets the plotting characteristics of the terminal
;
FILENAME=' '
;
FILENAME='6MAR3R' ;The file being used to generate the MGC data
;
SIZE=900          ;Number of data points in file
RAW=FLTARR(SIZE)
;
PRINT,'Reading(',FILENAME+'.DAT',')'
ON_IOERROR,LABEL1
OPENR,1,FILENAME+'.DAT' &READF,1,RAW
LABEL1: &CLOSE,1
;
;      Zeroing the baseline by averaging the points beyond 300
;
;
M=300 ;Maximum length of chromatogram
;
AVERAGE=(TOTAL(RAW(M:(SIZE-1)))/(SIZE-1-M))
RAW=RAW-AVERAGE
AVERAGE=(TOTAL(RAW(M:(SIZE-1)))/(SIZE-1-M))
RAW=RAW-AVERAGE
;
PRINT,'%%%%%%%%% Multiplexing the initial array
RAW %%%%%%%%%'
;
;      Select the number of injections desired
;
;
READ,'ENTER THE NUMBER OF INJECTIONS ',NU
READ,'Enter the noise amplitude ',MAXI
;
;      Variables used
;
;
T=FLTARR(NU)

```

```

SPACING=INTARR(NU)
;
MA=50 ;The maximum time, in seconds, between injections
MI=10 ;The minimum time between injections
;
SPACING=(FIX((MA-MI)*RANDOMU(SEED,NU)))+FIX(MI) ;The
random array that
;
; ranges from 10 to 50
;
TOTALDELAY=TOTAL(SPACING)
;
; Variables used
;
INDEX=SIZE+TOTALDELAY ;The total number of data points
TEMPLATE=FLTARR(INDEX)
;
FOR I=0,SIZE-1 DO TEMPLATE(I)=RAW(I) ;Transfers initial data to
;
; template file
;
; This loop creates the MGC output (noiseless)
;
AHA=0
;
FOR J=0,NU-1 DO BEGIN
    BOB=AHA+SPACING(J)
    FOR I=0,SIZE-1 DO
        TEMPLATE(I+BOB)=TEMPLATE(I+BOB)+RAW(I)
    AHA=AHA+SPACING(J)
ENDFOR
;
; Adding noise to the output
;
SPAC=INTARR(INDEX)
SPAC=(FIX((MAXI)*RANDOMU(S,INDEX))) ;The noise
SPAC=SPAC-(MAX(SPAC)-MIN(SPAC))/2
SPAC=2*SPAC
TEMPLATE=TEMPLATE+SPAC ;The noise added to the output
(simulated MGC)
PLOT,TEMPLATE

```

```

;PRINT,'The MGC output for',NU,'injections'
\;
AA=FLTARR(INDEX)
INDEXX=0
;
AA(0)=1
FOR I=0,NU-1 DO BEGIN
    INDEXX=INDEXX+SPACING(I)
    AA(INDEXX)=1
ENDFOR
;PLOT,AA
;PRINT,'This is the random injection sequence for',NU,' injections (input)'
;
FTEMP=FFT(TEMPLATE,-1)
TEMPLATE=FFT(FTEMP,1)
FAA=FFT(AA,-1)
CHROM1=FFT(FTEMP/FAA,1)/(INDEX)
PLOT,CHROM1
PRINT,'The Fourier Transform deconvolution (The chromatogram)'
;
;    Now plotting the chromatogram over a given range and
;    and reading desired intensities
;
FIND=' '
ANOTHER=' '
;
LOOK:
READ,'Enter the range of interest #,# ',FIRSTPT,LASTPT
PLOT,CHROM1,XRANGE=[FIRSTPT,LASTPT]
READ,'Put on lock (CAPITAL LETTERS REQUIRED) '
READ,'Do you want an intensity reading? (Y/N) ',FIND
IF (FIND EQ 'Y') THEN BEGIN
    READ,'Enter point of interest (Integer) ',POINT
    PRINT,'Intensity at ',POINT,'= ',CHROM1(POINT)
ENDIF
READ,'Do you want to see a different range? (Y/N) ',ANOTHER
IF (ANOTHER EQ 'Y') THEN GOTO,LOOK
;
PRINT,'^^^^^^^^^^Applying the first filtering technique^^^^^^^^^^'

```

```

;
CHECK:
AAGAIN:
;
;   Variables used
;
CHEC=' '
PEAK=FLTARR(SIZE)
PEAK1=FLTARR(SIZE)
CHROM1=FLTARR(INDEX)
TEMPLATE1=FLTARR(INDEX)
TEMPLATE2=FLTARR(INDEX)
;
READ,'Enter the peak height from the Fourier deconvolution (real) ',X
READ,'Enter the known peak height (real) ',Q
READ,'Enter the BEGINNING,END of the peak to be removed
',END1,END2
READ,'Enter its retention time ',RETENTION
READ,'Did you enter the correct information? (Y/N) ',CHEC
IF (CHEC EQ 'N') THEN GOTO,CHECK
;
RATIO=X/Q
;
PEAK(END1:END2)=RAW(END1:END2) ;Creates a smooth peakshape
for the
;
;           single peak MGC
PEAK=RATIO*PEAK ;Retains the peak height found by the FT
deconvolution
HAHA=SORT(PEAK) ;Maintains the selected
HABIG=HAHA(SIZE-1) ; retention time
FOR J=0,20 DO PEAK1(RETENTION-5+J)=PEAK(HABIG-5+J);
;
AHA=0
FOR J=0,NU-1 DO BEGIN
    BOB=AHA+SPACING(J)
    FOR I=0,SIZE-1 DO
        TEMPLATE1(I+BOB)=TEMPLATE1(I+BOB)+PEAK1(I)
    AHA=AHA+SPACING(J)
ENDFOR

```

```

;PLOT,TEMPLATE1
;
TEMPLATE2=TEMPLATE-TEMPLATE1
;PLOT,TEMPLATE2
;
FTEMP1=FFT(TEMPLATE2,-1)
TEMPLATE2=FFT(FTEMP1,1)
FAA=FFT(AA,-1)
CHROM2=FFT(FTEMP1/FAA,1)/(INDEX)
PLOT,CHROM2
;
;   Variables used
;
AFIND=' '
AANOTHER=' '
ELIM=' '
;
ALOOK:
READ,'Enter range of interest #,# ',AFIRSTPT,ALASTPT
PLOT,CHROM2,XRANGE=[AFIRSTPT,ALASTPT]
READ,'Do you want an intensity reading? (Y/N) ',AFIND
IF (AFIND EQ 'Y') THEN BEGIN
    READ,'Enter point of interest (integer) ',APOINT
    PRINT,'Intensity at ',APOINT,'=',CHROM2(APOINT)
ENDIF

READ,'Do you want to see a different range? (Y/N) ',AANOTHER
IF (AANOTHER EQ 'Y') THEN GOTO, ALOOK
;
TEMPLATE=TEMPLATE2
;
READ,'Do you want to remove another peak? (Y/N) ',ELIM
IF (ELIM EQ 'Y') THEN GOTO, AAGAIN
END
[EOB]

```