



University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations

Graduate School

2009

**PATTERN RECOGNITION INTEGRATED SENSING
METHODOLOGIES (PRISMS) IN PHARMACEUTICAL PROCESS
VALIDATION, REMOTE SENSING AND ASTROBIOLOGY**

Thaddaeus S. Hannel
University of Kentucky, tshannel@gmail.com

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Hannel, Thaddaeus S., "PATTERN RECOGNITION INTEGRATED SENSING METHODOLOGIES (PRISMS) IN PHARMACEUTICAL PROCESS VALIDATION, REMOTE SENSING AND ASTROBIOLOGY" (2009). *University of Kentucky Doctoral Dissertations*. 751.

https://uknowledge.uky.edu/gradschool_diss/751

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Thaddaeus S Hannel

The Graduate School

University of Kentucky

2009

PATTERN RECOGNITION INTEGRATED SENSING METHODOLOGIES (PRISMS) IN
PHARMACEUTICAL PROCESS VALIDATION, REMOTE SENSING AND
ASTROBIOLOGY

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Chemistry at the University of Kentucky

By
Thaddaeus S. Hannel

Lexington, Kentucky

Director: Dr. Robert A. Lodder, Professor of Chemistry

Lexington, Kentucky

2009

Copyright © Thaddaeus S. Hannel 2009

ABSTRACT OF DISSERTATION

PATTERN RECOGNITION INTEGRATED SENSING METHODOLOGIES (PRISMS) IN PHARMACEUTICAL PROCESS VALIDATION, REMOTE SENSING AND ASTROBIOLOGY

Modern analytical instrumentation is capable of creating enormous and complex volumes of data. Analysis of large data volumes are complicated by lengthy analysis time and high computational demand. Incorporating real-time analysis methods that are computationally efficient are desirable for modern analytical methods to be fully utilized. The use of modern instrumentation in on-line pharmaceutical process validation, remote sensing, and astrobiology applications requires real-time analysis methods that are computationally efficient.

Integrated sensing and processing (ISP) is a method for minimizing the data burden and sensing time of a system. ISP is accomplished through implementation of chemometric calculations in the physics of the spectroscopic sensor itself. In ISP, the measurements collected at the detector are weighted to directly correlate to the sample properties of interest. This method is especially useful for large and complex data sets. In this research, ISP is applied to acoustic resonance spectroscopy, near-infrared hyperspectral imaging and a novel solid state spectral imager. In each application ISP produced a clear advantage over the traditional sensing method.

The limitations of ISP must be addressed before it can become widely used. ISP is essentially a pattern recognition algorithm. Problems arise in pattern recognition when the pattern-recognition algorithm encounters a sample unlike any in the original calibration set. This is termed the false sample problem. To address the false sample problem the Bootstrap Error-Adjusted Single-Sample Technique (BEST, a non-parametric classification technique) was investigated. The BEST-ISP method utilizes a hashtable of normalized BEST points along an asymmetric probability density contour to estimate the BEST multidimensional standard deviation of a sample. The on-line application of the BEST method requires significantly less computation than the full algorithm allowing it to be utilized in real time as sample data is obtained. This research tests the hypothesis that a BEST-ISP metric can be used to detect false samples with *sensitivity* > 90% and *specificity* > 90% on categorical data.

KEYWORDS: Process Analytical Technology, Chemometrics, False Sample, Bootstrap
Error-Adjusted Single-Sample Technique, Integrated Sensing and
Processing

Thad Hannel

Student's Signature

November 24, 2009

Date

PATTERN RECOGNITION INTEGRATED SENSING METHODOLOGIES (PRISMS)
IN PHARMACEUTICAL PROCESS VALIDATION, REMOTE SENSING AND
ASTROBIOLOGY

By

Thaddaeus S. Hannel

Robert Lodder

Director of Dissertation

Mark Meier

Director of Graduate Studies

November 24, 2009

Date

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

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

Name

Date

DISSERTATION

Thaddaeus S. Hannel

The Graduate School
University of Kentucky

2009

PATTERN RECOGNITION INTEGRATED SENSING METHODOLOGIES (PRISMS)
IN PHARMACEUTICAL PROCESS VALIDATION, REMOTE SENSING AND
ASTROBIOLOGY

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Chemistry at the University of Kentucky

By
Thaddaeus S. Hannel

Lexington, Kentucky

Director: Dr. Robert A. Lodder, Professor of Chemistry

Lexington, Kentucky

2009

Copyright © Thaddaeus S. Hannel 2009

In Memory of Patricia Ann Hannel

Acknowledgements

I would like to express my gratitude to all those who have offered their support, guidance, and patience. To my wife and family, I cannot express how much your support and confidence has meant to me. Special thanks to my advisor, from whom I have learned that knowledge can be a helpful guide, but imagination is the key to understanding.

TABLE OF CONTENTS

Acknowledgements.....	iv
List of Tables	viii
List of Figures.....	ix
List of Files	xi
Section I: Background.....	1
Chapter One- Pattern Recognition, Integrated Sensing and Processing and Chemometrics.....	1
Introduction	1
Pattern Recognition	3
Integrated Sensing and Processing	7
Chemometrics	10
Chapter One Figures.....	17
Section II: Pharmaceutical Process Validation	20
Chapter Two-BEST-SCARS for Process Validation of Experimental Formulations of a Novel Diabetes Drug.....	22
Introduction	22
Theory.....	27
Materials and Methods	28
Results and Discussion	31
Conclusion.....	35
Chapter Two Tables.....	36
Chapter Two Figures	40
Section III: Remote Alcohol Sensing	47
Chapter Three- Surreptitious Remote Sensing of Blood Alcohol Concentration: Integrated Near-Infrared Spectral Imaging and Laser Speech Detection	50
Introduction	50
Materials and Methods	54
Results and Discussion	60
Conclusion.....	63
Chapter Three Figures	66

Chapter Four-Bootstrap Error-adjusted Single-sample Technique Molecular Factor Computing (BEST-MFC) Near Infrared Hyperspectral Imaging: Classification of False Samples	77
Introduction	77
Materials and Methods	79
Results and Discussion	81
Conclusion	83
Chapter Four Tables	85
Chapter Four Figures	88
Section IV: Astrobiology	93
Chapter 5- Complementary Integrated Sensing and Processing Solid State Spectroscopy (CRISP-SSSI) for Rapid Detection of Copper Sulfate	96
Introduction	96
Theory.....	99
Materials and Methods	101
Results and Discussion	103
Conclusion.....	104
Chapter Five Tables.....	106
Chapter Five Figures	107
Chapter Six- BEST-CRISP for Classification of Common Martian Minerals and Identification of False Samples	114
Introduction	114
Theory.....	119
Materials and Methods	122
Results and Discussion	124
Conclusion.....	128
Chapter Six Tables.....	130
Chapter Six Figures	131
Chapter Seven- Size Does Matter: Solid State Spectral Imager (SSSI) on a Chip and the Search for Extraterrestrial Life.....	141
Introduction	141
Micro- and Nanotechnologies	143

Communicating with the SSSI	146
NASA and Microtechnology	148
Conclusion	149
Chapter Seven Figures	150
Section V: Conclusion of Dissertation.....	152
Appendix A: MatLab Code.....	154
A.1 BEST-ISP Code.....	154
A.2 SSSI MatLab Code.....	162
Appendix B: Solid State Spectral Imager User Manual	194
B.1 Instillation	194
B.2 Operation.....	195
B.3 Troubleshooting	198
Appendix B Figures.....	200
References.....	206
Vita.....	220

List of Tables

Table #	Table Title	Page
Table 2.1	Percent sample composition	37
Table 2.2	Performance statistics for BEST classification	38
Table 2.3	Performance statistics for BEST-SCARS	39
Table 2.4	Average BEST-SCARS calculated MSDs for false samples	40
Table 4.1	Erroneous predictions	85
Table 4.2	Performance statistics for BEST classification	86
Table 4.3	Performance results of the BEST-MFC	87
Table 5.1	CRISP resolution example	105
Table 6.1	BEST-CRISP analysis of false samples	128

List of Figures

Figure #	Figure Title	Page
Figure 1.1	Illustration of synthetic bivariate data	18
Figure 1.2	BEST multidimensional standard deviation contour	19
Figure 1.3	Near-infrared correlation chart	20
Figure 2.1	Schematic representation of the ARS	41
Figure 2.2	SCARS illustration	42
Figure 2.3	Pure component spectra	43
Figure 2.4	XYZ Scatter Plot: PCs	44
Figure 2.5	XYZ Scatter Plot: Voltage	45
Figure 2.6	Mean BEST MSDs	46
Figure 2.7	XYZ Scatter Plot: Voltage with False Samples	47
Figure 3.1	Remote alcohol sensing	66
Figure 3.2	MFC illustration	67
Figure 3.3	Subjects Facial Profile	68
Figure 3.4	Comparison of MFC and PCA image calibration results	69
Figure 3.5	Comparison of PCA and MFC image calibration SEP values	70
Figure 3.6	Global calibration plots	71
Figure 3.7	Speech spectra from patient 5	72
Figure 3.8	Interferometer-recorded speech correlations	73
Figure 3.9	Interferometer-recorded speech SEP values	74
Figure 3.10	Microphone-recorded speech correlations	75
Figure 3.11	Microphone-recorded speech SEP values	76
Figure 4.1	Near-infrared correlation chart	88
Figure 4.2	Reflectance data for selected pixel regions	89
Figure 4.3	Estimated BEST MSD contour at 3 SDs	90
Figure 4.4	Estimated BEST MSD plot of subject 9	91
Figure 4.5	Estimated BEST standard deviation plot	92
Figure 5.1	Hadamard encoding description	106
Figure 5.2	SSSI prototype	107
Figure 5.3	CRISP method example	108
Figure 5.4	Mean transmission spectra	109
Figure 5.5	Cross validation plots	110
Figure 5.6	Positive pulse sequence duty cycles	111
Figure 5.7	Negative pulse sequence duty cycles	112
Figure 6.1	SSSI prototype robot	129
Figure 6.2	PC plot of simulated data	130
Figure 6.3	CRISP loading example	131
Figure 6.4	Positive and negative loading pulse sequences	132
Figure 6.5	Comparison of BEST-CRISP to the full BEST MSDs	133
Figure 6.6	PC plot of mineral data	134
Figure 6.7	Hadamard scan BEST MSDs	135
Figure 6.8	CRISP scan BEST MSDs	136
Figure 6.9	BEST-CRISP MSDs	137

Figure 6.10	Voltage plot of mineral and false samples	138
Figure 7.1	Hadamard/CRISP flow chart	149
Figure 7.2	CRISP example	150
Figure B.1	Diagram of the SSSI connection sequence	198
Figure B.2	Screen shot of the SSSI GUI	199
Figure B.3	SSSI GUI in Hadamard scanning mode	200
Figure B.4	Kiel uVision3 screen shot	201
Figure B.5	Kiel uVision3 error and warning messages	202
Figure B.6	Silicon Labs IDE software screen shot	203

	List of Files	
Figure #	File name	Size
Figure 1.1	Figure1.1.TIF	831 k bytes
Figure 1.2	Figure1.2.TIF	838 k bytes
Figure 1.3	Figure1.3.TIF	1,063 k bytes
Figure 2.1	Figure2.1.GIF	453 k bytes
Figure 2.2	Figure2.2.TIF	1,181 k bytes
Figure 2.3	Figure2.3.TIF	1,036 k bytes
Figure 2.4	Figure2.4.TIF	1,194 k bytes
Figure 2.5	Figure2.5.TIF	1,194 k bytes
Figure 2.6	Figure2.6.TIF	1,114 k bytes
Figure 2.7	Figure2.7.TIF	1,194 k bytes
Figure 3.1	Figure3.1.JPEG	210 k bytes
Figure 3.2	Figure3.2.JPEG	388 k bytes
Figure 3.3	Figure3.3.GIF	50 k bytes
Figure 3.4	Figure3.4.TIF	724 k bytes
Figure 3.5	Figure3.5.TIF	727 k bytes
Figure 3.6	Figure3.6.TIF	2,605 k bytes
Figure 3.7	Figure3.7.JPEG	183 k bytes
Figure 3.8	Figure3.8.TIF	800 k bytes
Figure 3.9	Figure3.9.TIF	783 k bytes
Figure 3.10	Figure3.10.TIF	801 k bytes
Figure 3.11	Figure3.11.TIF	782 k bytes
Figure 4.1	Figure4.1.TIF	1,063 k bytes
Figure 4.2	Figure4.2.TIF	1,827 k bytes
Figure 4.3	Figure4.3.TIF	1,276 k bytes
Figure 4.4	Figure4.4.TIF	1,017 k bytes
Figure 4.5	Figure4.5.TIF	1,276 k bytes
Figure 5.1	Figure5.1.GIF	9 k bytes
Figure 5.2	Figure5.2.TIF	1,181 k bytes
Figure 5.3	Figure5.3.GIF	17 k bytes
Figure 5.4	Figure5.4.TIF	2,232 k bytes
Figure 5.5	Figure5.5.TIF	161 k bytes
Figure 5.6	Figure5.6.TIF	1,296 k bytes
Figure 5.7	Figure5.7.TIF	355 k bytes
Figure 6.1	Figure6.1.TIF	17,955 k bytes
Figure 6.2	Figure6.2.TIF	1,036 k bytes
Figure 6.3	Figure6.3.GIF	17 k bytes
Figure 6.4	Figure6.4.TIF	16,276 k bytes
Figure 6.5	Figure6.5.TIF	664 k bytes
Figure 6.6	Figure6.6.TIF	1,036 k bytes
Figure 6.7	Figure6.7.TIF	1,113 k bytes
Figure 6.8	Figure6.8.TIF	1,113 k bytes
Figure 6.9	Figure6.9.TIF	1,113 k bytes
Figure 6.10	Figure6.10.TIF	2,311 k bytes
Figure 7.1	Figure7.1.JPEG	69 k bytes

Figure 7.2	Figure7.2.GIF	17 k bytes
Figure B.1	Figure B.1.TIF	661 k bytes
Figure B.2	Figure B.2.TIF	509 k bytes
Figure B.3	Figure B.3.TIF	640 k bytes
Figure B.4	Figure B.4.TIF	1,179 k bytes
Figure B.5	Figure B.5.TIF	565 k bytes
Figure B.6	Figure B.6.TIF	534 k bytes

Section I: Background

Chapter One- Pattern Recognition, Integrated Sensing and Processing and Chemometrics

Introduction

Modern analytical instrumentation is capable of collecting an enormous amount of data. Hyphenated and hypernated techniques are commonly employed and produce a wealth of information from a single sample¹⁻³. The large and diverse data volumes that are generated from modern techniques must be carefully analyzed to reveal meaningful and useful relationships among the data. This becomes a difficult task when real-time results are needed. Consider the combination of hyperspectral imaging and synthetic aperture radar (SAR) for military target identification⁴. The combining of these imaging techniques generates both spectral and spatial feature information of target scenes. The sensor output may comprise millions of pixels covering hundreds of discrete wavelength intensities. In a combat situation HSI alone is capable of producing greater than 10^{12} pixels/day which would require on the order of 10^{16} flops/day to process⁵. Taking into account that valuable target information may be hidden within the data, there is a need to rapidly reduce the data from physical fields to high level information.

Integrated Sensing and Processing (ISP) is a method of physically reducing high-dimensional data at the sensor to produce a low-dimensional representation of the underlying high-level information. ISP was initiated by the Defense Advanced Research

Projects Agency (DARPA) as a paradigm for the implementation of mathematical processes into sensor design⁶. There are a number of mathematical (chemometric) methods that are commonly applied to determine correlations within data sets. The application of ISP in the data collection process essentially weights the raw data according to a chemometric method. This weight function results in low-dimensional high-level data that resemble a mathematic computation. In this respect, ISP is essentially a pattern recognition algorithm in that it produces a mathematical computation that resembles a predefined pattern.

Modern microelectronics, optics, mechanical and computer technology facilitate the acquisition of high-quality and large data volumes. Today, the Analytical chemist typically spends more time analyzing data than on its actual acquisition. To this end, there is sufficient reason to explore methods which simplify the data analysis procedure. The following chapter is meant to serve as an introduction to pattern recognition and integrated sensing and processing. Within the following section the chemometric techniques used will also briefly be discussed.

Pattern Recognition

The ability to classify patterns is accomplished easily by humans yet instilling this in a machine becomes an exceedingly complex task and has yet to be fully realized. Pattern recognition (PR) is a field of science that has been evolving since the development of computer technology. There are innumerable applications of PR which range from common programs like automated bill payment software or speech and text recognition to the more extensive and elusive applications such as artificial intelligence. The Oxford English dictionary defines PR as “*the process by which a computer, the brain, etc., detects and identifies ordered structures in data or in visual images or other sensory stimuli*”⁷. While computers cannot be programmed to match the human ability to identify ordered structures they can be programmed to tackle statistical specific tasks. For example, Hyperspectral imaging (HSI) of aspirin tablets can easily be done on-line. Each HSI may contain information on hundreds of tablets moving down a process line. Imagine that for each image pixel the physical observations, such as aspirin concentration, of a single tablet are represented as a $1 \times n$ vector, \mathbf{x} where n represents the number of discrete wavelengths, and can be mapped onto a single point in n -dimensional space, \mathbf{X} . When the physical properties of multiple vectors are similar they lie close together in \mathbf{X} . Using *a priori* knowledge of the statistical distribution of aspirin tablets of specific concentrations, pattern recognition can be used to determine whether the tablets on the process line are within a defined concentration range. Therefore, the position of \mathbf{x} in \mathbf{X} -space may be used to make a prediction about the physical properties of an aspirin tablet. Figure 1.1 describes this process. For the pattern recognition routine to be complete a decision boundary in \mathbf{X} -space must be specified. The decision boundary is

identified according to a classification routine. In this work the bootstrap error-adjusted single-sample technique (BEST) is used to determine the class of objects in an n -dimensional space. The BEST creates a nonparametric boundary (contour) in multidimensional standard deviations (MSDs) around each cluster in X-space. Objects falling within a 3 MSD contour are considered to be from the same population. Figure 1.2 illustrates the use of BEST distances as a decision boundary for the pattern recognition algorithm.

With the use of PR in analytical sensors it has become possible to automate many measurement processes. The automation and real-time control of measurement processes is at the center of what is termed process analytical technology (PAT). PAT is a paradigm with the goal of understanding and controlling manufacturing processes⁸. Although PAT is a Food and Drug Administration (FDA) initiative the advantages associated with it extend to other manufacturing areas. It is widely recognized that on-line and at-line measurements are required for better process understanding and have the ability to save time and cost while producing a better quality product. A system that is capable of controlling or predicting a measurement through real-time control of the process is termed a dynamic data driven application system (DDDAS).

In a DDDAS data is incorporated in to the measurement process dynamically rather than statically. In this manner the predictions and measurements become more reliable. In the HSI example previously used, a DDDAS system would control the manufacturing process based on real-time feedback of tablet data. As manufacturing conditions change the DDDAS would have the ability to adjust in real-time. This is a process that is commonly applied to hurricane forecasting. Data from the US Air Force

53rd Weather Reconnaissance Squadron (“Hurricane Hunter”) aircraft is used to update computer models in real-time to make better predictions and storm tracking forecasts. The prediction accuracy based on the updated information has produced as much as a 30% increased in warnings and predictions⁹.

The advantages of PAT and DDDASs are evident however; they are difficult to implement because of the processing requirements associated with multivariate data. Another disadvantage is the use of multivariate data in prediction models. Techniques such as HSI can produce a large number of intensity values over a wide range of discrete wavelengths. Problems arise when including many dimensions into the classification model. This is known as the “curse of dimensionality” and there are many publications on the subject¹⁰⁻¹². A detailed examination of the subject is out of the scope of this paper. However, it is well understood that for a training set of size $m \times n$, where m is the number of samples and n , the number of predictor variables, m must be much larger than n . Generally it is taken that to estimate the sample distribution of m in n -dimensional space the ratio of samples to dimensions (predictor variables) must be greater than ten ($m/n > 10$)¹³. Because of this it is often necessary to use chemometric techniques to extract a small number of useful features from the data.

Feature extraction is an important aspect in a PR algorithm. The feature extraction process works to reduce the dimensionality of the PR space which has advantages in measurement cost and classification accuracy. Dimensionality reduction also aids in the visualization of the data, storage and retrieval, and noise removal. There are a number of data reduction techniques available such as principal component analysis (PCA)¹⁴⁻¹⁵, kernel PCA¹⁶, partial least squares (PLS)¹⁷, and Fisher linear discriminant

analysis¹⁸. Data reduction techniques function to reduce multivariate data to high-level information. This step is an essential part of the PR process. However, many of these techniques require a high computational cost. The purpose of the aforementioned statistical methods is to find a smaller subset of the multivariate data that accurately represents the physical properties of the sample. This generally involves the translating and rotation of the data to a new coordinate system. This can be done by calculating the eigenvectors and eigenvalues of the data covariance matrix. Calculation of the data covariance matrix cannot always be done at an efficient rate to handle large data feeds such as those from on-line imaging. The development of rapid and real-time data reduction techniques will improve future PR systems and facilitate the use of PAT sensor systems throughout the manufacturing industry. One such method of reducing data to high-level useful information is by integrating the sensing and processing directly into the detector.

Integrated Sensing and Processing

An alternative approach to performing factorization techniques post data acquisition is to perform them during the acquisition process. Integrated sensing and processing (ISP) is a method that incorporates information about the sample directly into the physical design of the data collection process. ISP is an alternative approach to traditional factorization techniques. In traditional methods sampling is performed to obtain data which is then analyzed to reveal information about the sample. The statistics gained from the analysis can be applied to future data sets but computation will always follow sampling. The computational processes following sampling in traditional methods are applied with ISP during the sampling process by weighting the raw data at the sensor, making computation following data acquisition unnecessary. This works well for situations where there is more than enough time for calibration before data collection, but limited time after. For example, consider the PAT initiative in the pharmaceutical industry. An ideal PAT for on-line processes must be nondestructive and have the ability to make accurate and rapid measurements. It would be very difficult to analyze hundreds of terabytes of HSI data from tablets moving down a process line in real time. As imaging technology becomes smaller and cheaper, the ratio of “pixels to pupils” is moving toward infinity. It is certainly possible with today’s technology to collect data for 100% of the pharmaceutical tablets on a production line. It is analyzing all of that data and acting on it that is problematic. An ISP approach to process monitoring would work well under these circumstances where high sample throughput is required.

ISP is well suited for optical and acoustic spectroscopy. In optical spectroscopy ISP is achieved by weighting the optical spectrum at the detector to encode high-level

information about the analyte. In acoustic resonance spectroscopy ISP is accomplished by weighting the acoustic excitation waveform to encode high-order information at the detector. There are numerous approaches to ISP in both acoustics and optical spectroscopy. Three ISP methods are discussed in this paper and the theoretical approaches can be found in their respective sections. The commonality between each method is the use of principal components in the algorithm design. The implementation of an ISP sensor into the data collection design has the advantages of a reduced computational cost, increased throughput and signal-to-noise.

The advantages of ISP as a PAT in sensing processes are evident. However, the limitations of ISP must be addressed before it can become widely used. ISP is essentially a pattern recognition algorithm that can be performed both optically and acoustically. Problems arise in pattern recognition when the pattern-recognition algorithm encounters a sample unlike any in the original calibration set. This is termed the false sample problem. The false sample problem is only an issue if the calibration set used to select the ISP method contained every substance in the known universe. Figure 1.3 illustrates the absorption regions from overtones and combinations of common functional groups in the near-IR region of the electromagnetic spectrum. Consider the use of an ISP algorithm in near-infrared (IR) hyperspectral imaging for ethanol detection. If a sample contains nothing but ethanol, theoretically a filter could be used to weight all light transmitted to the detector except that of ethanol. An ISP filter for hyperspectral imaging could be composed from molecules containing all functional groups except $-OH$. All light would be blocked (weighted) by the filter except light that reveals ethanol concentration. The ISP filter would be immune to false samples as long as functional groups absorbed light

completely and only where indicated in a spectra-structure correlation chart. However, in real life, everything absorbs light at all wavelengths in the near-IR. Different chemicals just absorb light in slightly different amounts at all wavelengths. It is this property of absorbance that creates a need for multivariate analysis and can lead to false-sample problems.

For example, consider contamination of pharmaceuticals. Millions of pills are recalled each year at great expense. In 2006, 500 mg acetaminophen tablets were contaminated with metal wire leading to the recall of 11 million bottles¹⁹. Metal wires act like perfect electrical conductors at near-IR wavelengths and thus have no distinguishing spectral features. Even if a PAT HSI system had been in place to analyze 100 % of the tablets, it is possible that the contaminant would have been predicted as a reasonable concentration of the drug. Calibration of an ISP algorithm would not have considered metal in the training set. The false sample problem is further exacerbated when the foreign object has a similar spectrum to the analyte (isospectrality).

Integrated sensing and processing strategies for optical and acoustic spectroscopy have potential for use in many areas because of the real-time processing power. However, it is impossible to integrate sensing and processing to encompass all conceivable contaminants. For these reasons a fast, ISP method of false sample detection is needed.

Chemometrics

According to the International Chemometrics Society (ICS) chemometrics is “*the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods*”. With the advancements of microelectronics and computer technology today, the analyst is capable of acquiring high quality information with relative ease. These advancements have also led to better understanding of the basic theories and methodologies associated with modern analytical chemistry. The field of chemometrics has led to a wide range of new theories and algorithms for manipulating chemical data as well. Many chemometric theories and applications are routinely applied in all areas of chemistry from pharmaceutical sciences to chemical engineering. In this work there are several commonly applied chemometric techniques that are used such as multiplicative scatter correction, cubic spline smoothing and multiple linear regression (MLR) that will not be presented in this section. For information on these techniques the reader is directed to several source materials²⁰⁻²². Two techniques which are at the center of this work will, however be covered in this section. Principal component analysis (PCA) is a data reduction technique that is used in this work and is what the ISP techniques are based on. The bootstrap error adjusted single sample technique (BEST) is used throughout for classification of data clusters and as for false sample detection and thus will be presented here as well. The performance metrics for evaluating the classification models will also be presented.

Principal Component Analysis (PCA). PCA is one of the most commonly used tools for preprocessing multivariate data^{14-15,23}. PCs are calculated according to the singular value decomposition (SVD). The SVD is considered to be a very important computational tool

because of its properties such as its use in solving ill-conditioned systems²⁴. This is especially useful for analyzing physical data because in most situations the physical observations will lead to ill-conditioned systems. Ill-conditioned systems are problematic due to the fact that very small changes in observed data can cause very large changes in calculated answers. These changes can have an enormous effect on predictions by causing very large errors. For example, consider the equation $Ax = b$

$$\begin{bmatrix} 400 & -201 \\ -800 & 401 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 100 \\ -400 \end{bmatrix}$$

The system is easily solved and the solutions for x_1 and x_2 are found to be 101 and 200, respectively. Introducing a small change to one number in the system can have an enormous effect on the answer. For example, $a_{11} = 401$.

$$\begin{bmatrix} 401 & -201 \\ -800 & 401 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 100 \\ -400 \end{bmatrix}$$

Solving this system, the solutions for x_1 and x_2 are now calculated to be -40300 and -80400, respectively. The large error in the answer is evident.

SVD begins by the factorization of an ($m \times n$) matrix \mathbf{A} (for $n > m$) according to Equation 1.1.

$$A = VSU^T \tag{1.1}$$

Where \mathbf{V} is an $m \times m$ orthogonal matrix, \mathbf{S} is an $m \times m$ matrix of positive and real *singular values* of \mathbf{A} and are ranked according to their variance such that $s_{11} \geq s_{22} \geq s_{mm}$, and \mathbf{U}^T is an orthogonal $n \times n$ matrix. The standard principal component solution,

\mathbf{D} is formed from the combination of \mathbf{V} and \mathbf{S} . Two important aspects of PCA can be expressed by equation 1.2.

$$A^{(k)} = \sum_{i=1}^k d_i u_i \quad 1.2$$

Where d_i and u_i are the i^{th} column vectors of the \mathbf{D} and \mathbf{U} matrices and $A^{(k)}$ is the closest rank- k matrix to \mathbf{A} such that $A^{(k)}$ minimizes the sum of the squares of the difference of the elements of \mathbf{A} and $A^{(k)}$, $(\sum_{ij} |a_{ij} - a_{ij}^k|^2)$. One, small singular values have little effect on the summation while large singular values contribute significantly to the sum, and two, when small singular values are omitted from the summation the corresponding rank is lower. It is important to note that if a variable in \mathbf{A} has a much larger variance than the rest it will dominate the first PC. To avoid this situation each variable in the raw data is usually standardized to zero mean and unit variance. When the raw data is normalized there is a direct correspondence between the covariance matrix of \mathbf{A} and the principal components, \mathbf{D} calculated from \mathbf{S} which are ordered according to the variance of the data set. The PC loadings, \mathbf{U} are calculated such that the corresponding PCs are not correlated. PCA is equivalent to a rotation of axis such that the first PC contains the largest amount of variation the second PC contains the next greatest and so on until all of the variation is accounted for.

PCA is an excellent data reduction technique for large data set. In most cases only a few PCs are needed to represent the variation in the data set. In a pure two component system the variation of the components could theoretically be captured by the first two PCs assuming the components were not correlated. However in practice this will not always be the case. Interferences, detector noise, drift and other inconsistencies

will no doubt make the number of PCs needed to describe the variation greater than that of the individual components. It is generally accepted however, that the number of PCs should be less than half of the number of training set samples used in the analysis.

Bootstrap Error-adjusted Single-sample Technique (BEST). The BEST is a nonparametric method for estimating the distribution of a sample²⁵⁻²⁹. The BEST metric provides advantages over other methods because it does *not* assume equal covariance for all spectral groups and that each group is drawn from a normally distributed population. The BEST uses the method of bootstrapping for statistical inference of the true sample distribution from a finite sample set³⁰⁻³³. The BEST method begins by acquiring an $m \times n$ training set sample \mathbf{T} , where m corresponds to the number of sample measurements and n corresponds to independent variables (e.g., reflectance or absorbance). The training set \mathbf{T} , is a representative sample of the population \mathbf{P} , and is chosen to describe the population variation in the universe of possible samples. A discrete realization (\mathbf{P}^*), of \mathbf{P} is calculated with the bootstrap operation $\kappa(\mathbf{T})$. The bootstrap operation begins by drawing j random samples for $j = (1, 2, \dots, m)$ from \mathbf{T} , with replacement. Bootstrap sample sets $\mathbf{B}_{(s)}$, are constructed from the random selections of \mathbf{T} and make up the i^{th} rows of \mathbf{B} for $i = (1, 2, \dots, k)$ according to:

$$b_i = \sum_{i=1}^m b_{(s)i} / m \quad 1.3$$

And the center \mathbf{C} , of the bootstrap distribution is calculated according to:

$$c_i = \sum_{i=1}^k b_{ij} / k \quad 1.4$$

The $k \times n$ bootstrap distribution \mathbf{B} , is the Monte Carlo approximation to the distribution. While there is no specified criterion for the number of bootstraps (k) used in the approximation, the number should be much larger than the number of samples. In this work 2000 bootstraps are used in all BEST calculations unless otherwise specified.

Once the bootstrap distribution has been calculated the analysis of a test sample can be completed. For a test sample \mathbf{X} ($l \times n$), the operation $\Psi(T, B, X, C)$ is used to calculate the MSD of \mathbf{X} from the center, \mathbf{C} of the distribution. The standard deviation, σ in the direction of \mathbf{X} is calculated according to Equation 1.5.

$$\left\{ \sigma \left| \frac{\int_0^\sigma \int_R P^* \rightarrow \overline{CX}}{\int_R P^* \rightarrow \overline{CX}} = 0.68 \right. \right\} \quad 1.5$$

The upper and lower confidence limits are skew corrected based on the comparison of the expected value of \mathbf{P} , $C = E(P)$ and the median value of \mathbf{T} , $C_T = med(T)$. The correction is projected onto the hyperline connecting \mathbf{C} and \mathbf{X} according to:

$$(C - C_T) \rightarrow \overline{CX} \quad 1.6$$

The result of the corrected projection is a skewed MSD in the positive and negative direction of \mathbf{X} according to Equations 1.7 and 1.8:

$$\left\{ \begin{array}{l} \rightarrow \\ +\sigma \end{array} \left| \frac{\int_0^{+\sigma} \int_R P^* \rightarrow \overline{CX}}{\int_R P^* \rightarrow \overline{CX}} = 0.34 \right. \right\} \quad 1.7$$

$$\left\{ \begin{array}{l} \rightarrow \\ -\sigma \end{array} \left| \frac{\int_0^{-\sigma} \int_R P^* \rightarrow \overline{CX}}{\int_R P^* \rightarrow \overline{CX}} = 0.34 \right. \right\} \quad 1.8$$

The asymmetric confidence limits make no assumption about size, shape or orientation of the spectral cluster in hyperspace.

Performance Metrics. In every analytical measurement there exists some degree of error. While the analyst will make every effort to minimize this error it cannot be avoided. To this end, there must always be an estimation of the uncertainty in the data if the results are to be accepted. In this work regression models and classification models were evaluated using the performance metrics described below.

Regression Models. For regression models leave-one-out cross validation (LOOCV) was performed. LOOCV, starts with $n + 1$ samples, the last n samples are used to calculate a calibration, which is then used to predict the value of the first sample. The first sample is then put back and the second sample is left out to form the second calibration, which is then used to predict the second sample. The process continues until each sample has been left out of the calibration and predicted once. Predictions from LOOCV regression were assessed using the correlation coefficient (R^2), the standard error of estimate (SEE) and standard error of prediction (SEP). The F test (a ratio of SEE^2/SEP^2) is employed to establish the equivalence of the calibration and validation errors for cross validation. The null hypothesis for the test was that the calibration errors and the validation errors were the same (i.e., F is close to 1). If the probability returned by the F test is high (typically greater than 0.05), the null hypothesis is accepted and the calibration is deemed acceptable. If the F probability is less than 0.05, potential problems in the sample selection or calibration method (e.g., over-fitting) must be investigated.

Classification Models. In classification models a set of samples must be assigned a class. This is generally referred to as a binary classification because the samples are either of the class (sample = 1) or not (sample = 0). Accuracy, precision, sensitivity and specificity were used to determine the performance of classification models and were calculated according to:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad 1.9$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad 1.10$$

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \quad 1.11$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \quad 1.12$$

Where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. The accuracy describes the overall correctness of the classification, precision describes the effectiveness of class assignments, sensitivity describes the effectiveness of the model for samples correctly identified with respect to the correct class and specificity describes the effectiveness of the model for samples successfully classified with respect to incorrect class. In other words sensitivity and specificity can be thought of as correct classes identified as correct and incorrect classes identified as incorrect, respectively.

Chapter One Figures

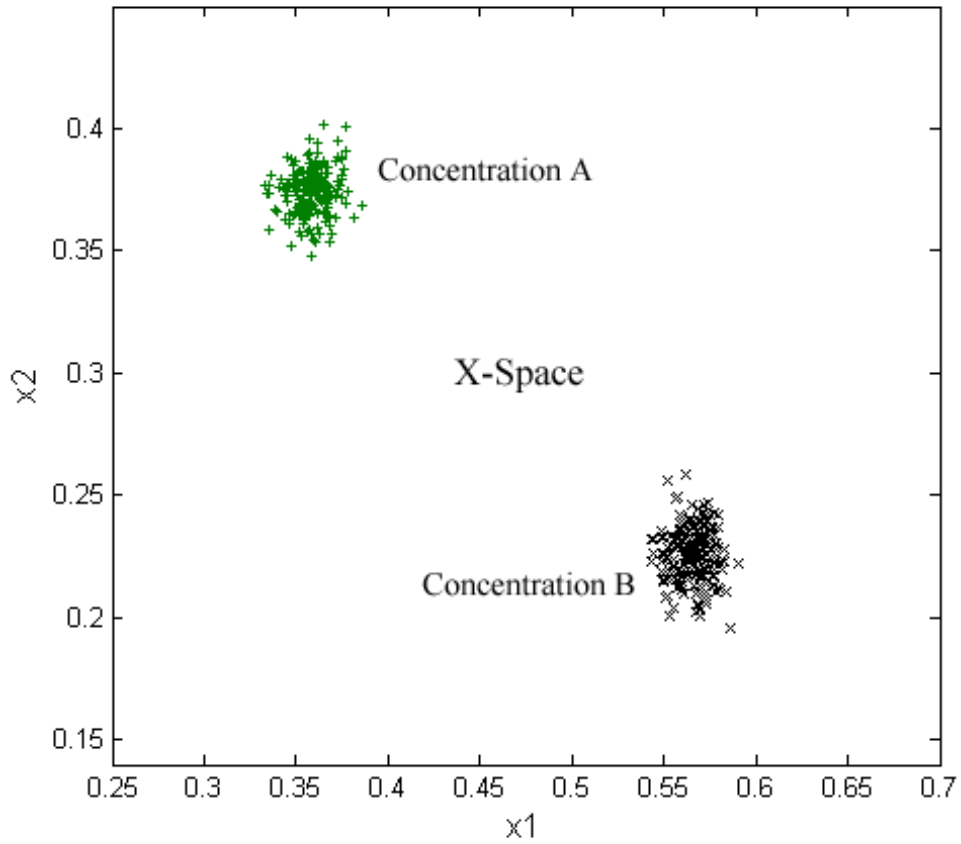


Figure 1.1: Illustration of synthetic bivariate data mapped into X-space where x_1 and x_2 are the measured values of two sample groups (represented by '+' and 'x'). The position of a data point in X-space can be used to determine the physical properties, such as concentration of the sample.

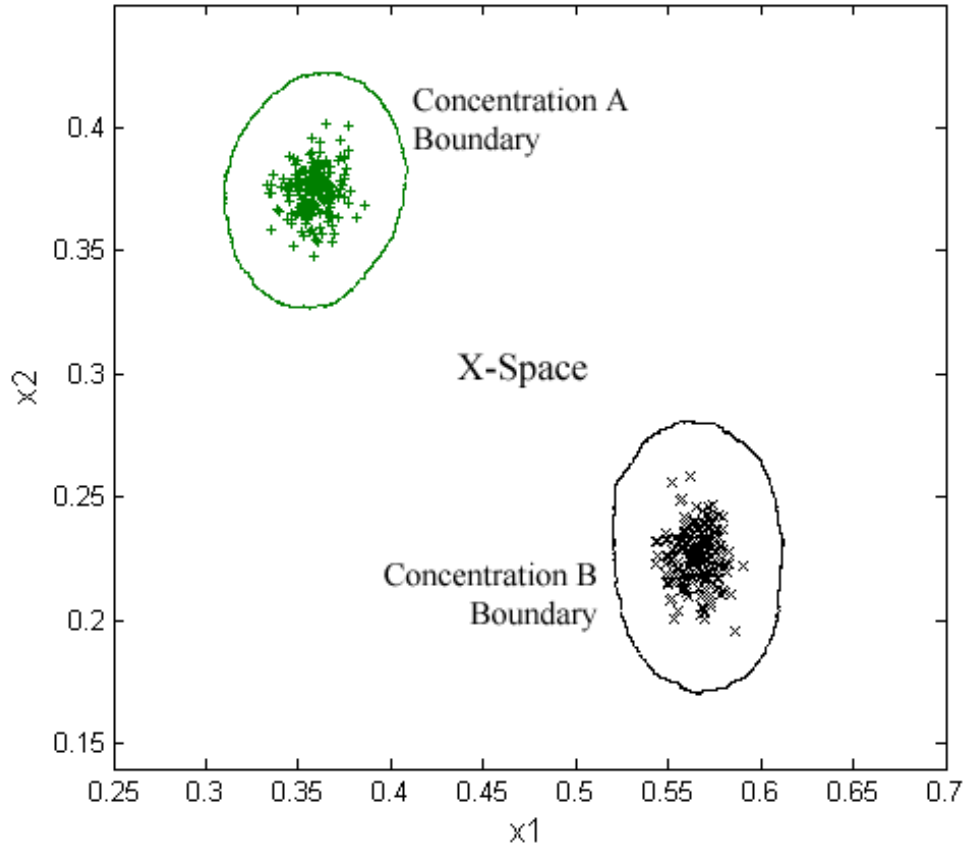


Figure 1.2: Illustration of synthetic bivariate data mapped into X-space where x_1 and x_2 are the measured values of two sample groups (represented by '+' and 'x'). A BEST multidimensional standard deviation contour is used to determine if samples in X-space belong to one of the two concentration groups.

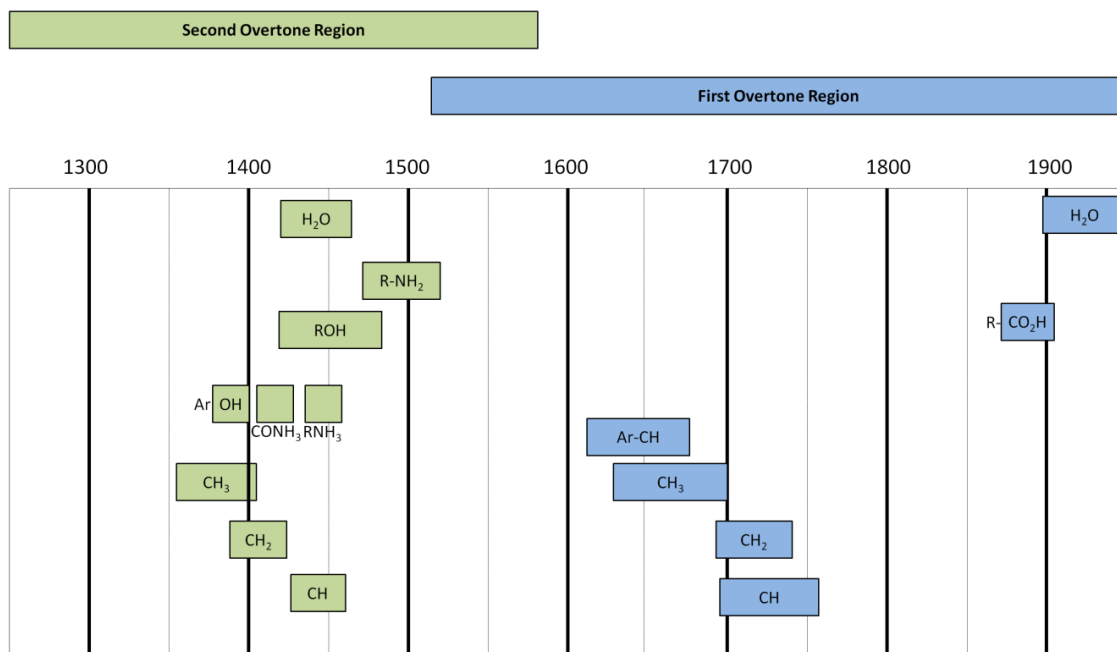


Figure 1.3: Near-infrared correlation chart for common functional groups. This chart is only an approximate guide for near-IR band assignment of functional groups.

Section II: Pharmaceutical Process Validation

Process validation encompasses the whole drug development process, with the end goal of ensuring that a product continually meets its predetermined requirements, and is required for all pharmaceutical products under Section 501(a)(2)(B) of the Federal Food, Drug, and Cosmetic Act. Process validation can be broken down into many different steps taking place over the lifecycle of a product. The FDA draft for process validation principles and practices describes process validation as taking place in three steps: process design, qualification, and continued process verification³⁴. In process design, the sensor system used to measure the product plays a role in validation. Product monitoring through sensor data can be used to establish a control strategy for manufacturing, thus the sensor also has a role in process qualification. The International Conference on Harmonisation (ICH) guidelines for pharmaceutical quality systems (ICH Q10) were designed to enhance current good manufacturing processes (cGMP) through better understanding of the pharmaceutical development cycle³⁵. One aspect of this is to provide feedback for process performance and product quality utilizing science and risk-based processes. While there are current protocols in place to monitor process performance and product quality, none are fool proof. Many products are recalled every year at great expense to the manufacturer. In June 2009, for example, Barr Laboratories, Inc. issued a recall of Dextroamphetamine/Amphetamine 20mg tablets due to an increase in the amount of the API in some tablets³⁶. In this case, the recall may have been avoided if better process monitoring were employed. The current trend toward better science and risk-based process understanding and control can be seen in the FDA's process analytical technology (PAT) initiative⁸. The goal of PAT is to build quality into the design of the

manufacturing process. In regards to the sensor system an ideal PAT would have the ability to make rapid, nondestructive, and accurate measurements. A PAT sensing system could have constant feedback and have the ability to continually control the manufacturing process.

Chapter Two-BEST-SCARS for Process Validation of Experimental Formulations of a Novel Diabetes Drug

Introduction

Diabetes. Diabetes is a chronic illness that affects more than 171 million people worldwide and is expected to double in incidence by 2030³⁷. Although the number of people affected by diabetes is much larger in developing countries, the cost of diabetes in developed countries is still very high. The total cost of diabetes in the United States in 2007 was estimated at \$174 billion³⁸. Of the total cost of diabetes, the direct medical cost associated with those diagnosed with the illness was \$116 billion³⁸. The U.S. Medicare program spent \$1.3 billion on diabetes-related hospital costs in 2001³⁹, illustrating that the cost of diabetes is not only high for those afflicted with the illness. Of those with diabetes 90% are diagnosed with Type II diabetes⁴⁰. Commonly referred to as adult onset diabetes, Type II diabetes usually occurs in overweight middle aged adults. Although manageable, Type II diabetes can cause many complications including heart attack, stroke, kidney disease, blindness, and even death. Type II diabetes occurs when the body can no longer effectively use insulin. The result of insulin inefficiency is a buildup of blood glucose levels measured as A1c (or HbA1c) levels. The United Kingdom Prospective Diabetes Study (UKPDS) indicated that aggressive treatment leading to the reduction of A1c levels in type II diabetes patients can significantly reduce the risk of microvascular disease⁴¹. However, aggressive treatment does not lower the risk of cardiovascular complications⁴². For most diabetes patients lifestyle changes (diet and

exercise) can make all the difference. However, as the number of diagnosis and deaths increase each year it is evident that the development of drugs to treat diseases such as diabetes is needed.

Novel Treatments. D-Tagatose (tag) is a rare but naturally occurring hexose bulk sweetener found in heated dairy products. There is interest in developing tag as an antidiabetic and obesity control drug and, it is currently in phase III clinical trials⁴³⁻⁴⁴. A recent study indicated that substitution of tag for sucrose as a source of carbohydrate in the diets of mice lowers the incident of obesity, hyperglycemia, and coronary artery disease⁴⁵. In a phase II clinical trial designed to determine the smallest dose of tag with a biological effect, tag was still shown to lower A1c levels in patients by as much as 0.3%⁴⁶. The ability to lower blood glucose levels and protect against cardiovascular diseases makes tag an excellent candidate drug for diabetes control.

Resveratrol (RSV; 3,5,4'-trihydroxystilbene) is most notably known for its association with the French paradox (high calorie diet and low incident of coronary heart disease). RSV is a natural component of red grapes and as such can also be found in wine. Studies have shown RSV to improve health and longevity of mice on high calorie diets⁴⁷. RSV is also known to possess a number of other pharmacotherapy effects⁴⁸⁻⁵¹, some of which make it a suitable choice for antidiabetic therapies. In a 2006 study, RSV was shown to lower instances of atherosclerosis and liver damage in diabetic mice⁵². RSV has also been demonstrated to have insulin-like effects while not overworking pancreatic β -cells, both important properties in the treatment of type 2 diabetes⁵³⁻⁵⁴. Pterostilbene (Stilbene; *trans*-3,5-dimethoxy-4'-hydroxystilbene) is chemically related to RSV and possesses many of the same pharmacotherapeutic properties. Stilbene can be

found naturally in berries and in the heart wood of the Indian Kino tree, which is known as a medicinal plant in India. Stilbene has been shown to decrease blood glucose and raise plasma insulin levels in diabetic rats⁵⁵⁻⁵⁶. The properties of the aforementioned drugs provide a basis for their use in the treatment of diabetes. The development of experimental drugs requires them to undergo vigorous scrutiny costing millions of dollars and taking many years before they can be approved for use. The validation of the final product is an important aspect of process validation and begins early in the development of new drugs.

Acoustic Resonance Spectroscopy. Acoustic resonance spectroscopy (ARS) is one potential PAT sensor. ARS is a rapid and nondestructive method for sample identification. ARS has been previously used in the analysis of pharmaceutical tablets⁵⁷⁻⁵⁹, powder compacts⁶⁰, liquids⁶¹, and semi-solids⁶². Figure 2.1 is a schematic representation of the ARS used in this research. The ARS comprises piezoelectric transducers (PZT) and a quartz rod which is kept in mechanical contact with a sample. Acoustic radiation is transmitted through the quartz rod via a transmitting PZT. A second PZT in the sample holder transmits acoustic radiation directly into the sample which constructively increases the resonance energy transmitted through the sample. Constructive and deconstructive interferences between acoustic waves in resonance with the sample and rod form the unique waveform that is recorded at the receiving PZT. Acoustic spectra generated by the ARS contain thousands of data points. For example, analysis of a sample at a sample rate of 44.1 kHz over 5 seconds generates 220,500 data points per recorded spectrum. The time-based acoustic spectra data are converted into the frequency domain using the Fourier transform (FT). The frequency spectrum of samples

obtained with the ARS may comprise a wide frequency range. The data used in this study comprised a frequency range between 0.020 and 20 kHz. Chemometric analysis of the Fourier transformed acoustic resonance (FTAR) data is used to determine the physical characteristics or identity of the sample. As a PAT sensor the ARS has the ability to gather data at fast speeds (milliseconds). However, performing the necessary chemometric calculations on acoustic data for qualification or quantification can be tedious and time consuming.

Sequencing computational acoustic resonance spectroscopy (SCARS), a method of integrated sensing and processing, is a new approach to FTAR spectroscopy. ISP has previously been applied to optical⁶³⁻⁶⁶ and acoustic^{59, 62} spectroscopies. In SCARS chemometric calculations are integrated into the detection process. In this method the excitation energy transmitted to the sample is weighted such that the signal at the detector corresponds to the physical characteristics or identity of the sample. Principal component analysis (PCA) is a commonly employed chemometric method for reducing large data sets into more manageable volumes¹⁵. In PCA the relationships and patterns within the data are more easily recognized. PCs are calculated such that the first contains the greatest amount of variation in the data, the second contains the next greatest (orthogonal to the first), and so on until all the variation has been accounted for. Thus, the relationships in FTARS data can be recognized with a much smaller number of PCs than with thousands of intensity values. Spectral *scores* are generated in PCA by weighting each signal, a in a narrow bandpass, at each frequency, f with a coefficient X over n frequencies.

$$score = X_1 a_{f_1} + X_2 a_{f_2} + X_3 a_{f_3} + \dots + X_n a_{f_n} \quad (2.1)$$

In SCARS an acoustic representation of the PC factors (X in Eq. 2.1) is constructed to weight the signal (a in Eq. 2.1) over a defined frequency range ($f_1 \dots f_n$ in Eq. 2.1). Figure 2.2 illustrates this process. The computational pulse sequences are transmitted through a sample resulting in an integrated signal, V_{score} that corresponds directly to a PC score according to equation 2.2.

$$V_{score} = \sum_j^J X_j a_{fj} \quad (2.2)$$

Where X_j is the loading at wavelength j , a_{ij} is the spectral amplitude of the i^{th} mixture at frequency j , and j is an index over J frequencies. In this research the largest 1, 10, and 100 PC factors (loadings) were used.

The computational pulse sequences generated for SCARS describe only the variations found in a specific training set. In this respect SCARS is essentially a pattern recognition algorithm designed to identify substances from a defined calibration set. Problems can arise when the pattern recognition algorithm encounters a sample outside of its predefined calibration range. This is termed the false sample problem. The intensities and peak frequencies measured with the ARS respond to both the intensive and extensive properties of a sample. All samples absorb and transmit acoustic energy at different frequencies. Because of these properties, substances outside of the SCARS training set can produce a false sample situation. Consider the Barr Laboratories, Inc. super-potent Dextroamphetamine/Amphetamine 20mg tablets³⁶. Calibration of a SCARS algorithm used to monitor tablet concentration may not have considered such a wide range of API concentrations. It is possible that the AR signals generated from contaminated tablets could be read as normal concentrations. It is impossible to integrate sensing and

processing to encompass all conceivable contaminants. For this reason, a fast ISP method of false sample detection is needed.

BEST-SCARS. BEST is a classification method based on the bootstrapping technique^{30, 33}. In the BEST algorithm, multidimensional standard deviations (MSDs) of a sample population are estimated without making assumptions about the distribution of the population^{25-26, 67}. The BEST metric provides advantages over other methods such as the Mahalanobis metric, because it does *not* assume equal covariance for all spectral groups and that each group is drawn from a normally distributed population. The BEST-SCARS method is capable of calculating BEST MSDs at speeds faster than the full BEST algorithm, giving BEST-SCARS potential as a PAT sensor. In BEST-SCARS standardized distances are generated from a SCARS training set. The distances and centers of each training set distribution are stored in a hash table for on-line use. On-line, estimated BEST distances are calculated by centering the test spectrum to each stored center and comparing its value to the distances in the hash table. With BEST-SCARS method rapid determination of group classification and identification of false samples is possible.

Theory

The voltage scores acquired from a SCARS training set are used to construct a hash table of distances. The hash table is constructed by first calculating the bootstrap distribution for each group in the training set. In this research, 2000 bootstraps replications were calculated for each training set group. BEST MSDs are calculated for each bootstrap

distribution using the BEST method. The ratio of a Euclidian distance to a BEST distance is calculated for each bootstrap according to Equation 2.3.

$$R = \frac{d(C,B)}{BESTSD} \quad (2.3)$$

Where $d(C, B)$ is the Euclidian distance from the center C , $(1 \times n)$ of the bootstrap distribution to the bootstrap point, B $(m \times n)$, $BESTSD$ $(m \times n)$ is a vector of BEST standard deviations for each bootstrap replicate, B and R $(m \times 1)$ is vector of ratios for each bootstrap. A matrix of normalized bootstraps, B_N $(m \times n)$ are calculated and stored in a hash table according to Equation 2.4.

$$B_N = C + [R * (B - C)] / \|B - C\| \quad (2.4)$$

Each new voltage spectrum, V generated from a SCARS scan is centered to each hash table distribution and normalized according to Equation 2.5.

$$V_N = V - C / \|V - C\| \quad (2.5)$$

The hash table is searched for the bootstrap point B_N , closest to the voltage spectrum, V_N by calculating the dot product of V_N and B_N . The largest dot product of B_N and V_N indicates the bootstrap B_N , closest to the test sample V . The BEST SD of V is then estimated according to Equation 2.6.

$$BESTSD = \frac{d(C,V)}{d(B,V)} \quad (2.6)$$

Where $d(C, V)$ and $d(B, V)$ are Euclidian distances.

Materials and Methods

Mixture Preparation. Samples consisted of combinations of D-tagatose (Spherix Inc.), resveratrol (Source Naturals, Inc., 40mg), and Pterostilbene (VDF FutureCeuticals, Inc., 0.1mg). The samples were ground in a mortar and pestle into a fine powder and hand-blended so that the concentration of each constituent in a sample mixture was a combination of 0, 12.5, 33.3, 50, 75, and 100% (see Table 1). A total of 9 mixture combinations were made. 500 mg of the powder mixture was combined with 50 mg each of casting resin (Bisphenol A/Epichlorohydrin) and polyamide curing agent (Versamid 140) (Firefox Enterprises, Pocatello, ID). The resulting paste was massed and hand-pressed to yield a 450mg pellet. The samples were allowed to cure for 48 hours prior to scanning. False samples were made in the same manner. Three false samples were constructed containing 100% D-glucose (Fisher Scientific), a third mixture each of D-glucose, RSV and Stilbene, and 100% vitamin B-12 (Spring Valley, 2,000 mcg).

ARS Data Collection. A spring-loaded scale was used to keep samples in mechanical contact with the vertex of the quartz rod. The scale was adjusted so that there was a constant 200g of pressure between the sample and rod. White noise excitation was generated in MatLab and transferred to a MP3 player (Zoom Corp., H4). The audio output of the MP3 player was used to transmit the excitation signal to the transmitting PZTs. The audio output of the MP3 player was adjusted so that the signal sent to the ARS was at the maximum level. Data were acquired over 10 s with a sample rate of 44.1kHz using an external sound card (MobilePre USB, M-Audio), which allowed for amplification of the PZT output. All processing and analysis was performed in MatLab R2008a (The Mathworks Company, Natick, MA, USA). Analysis of ARS data was done in the frequency domain by performing the Fourier Transform (FT) on each time domain

spectrum. This FT resulted in frequency spectra that contained 22,050 intensity values. The frequency spectra contained a large resonance peak over a frequency range of 3.1 to 4.4 kHz that was subsequently selected for the analysis. PCA was performed on the z-scored frequency data over the defined range. The BEST method was used to determine MDSs between sample clusters. Three PCs were used to classify the sample set.

Acoustic Sequence Construction. Acoustic sequences were generated using frequencies corresponding to loading values generated from PCA. Positive and negative loading data were separated and the frequencies corresponding to the largest 1, 10, and 100 loading values were used to create acoustic sequences. Each acoustic sequence was 500ms in length. A total of three PCs separated into their positive and negative counterparts yielded 6 acoustic sequences for each 1, 10, and 100 models. SCARS excitations for each frequency model were 3 s in length and comprised three positive sequences followed by three negative sequences. Each 1, 10, and 100 frequency acoustic sequence models were placed in succession. The total length of the final acoustic excitation was 9 s.

BEST-SCARS Data Collection. SCARS data was acquired in the same manner as FTARS data. The average detector voltage for each negative acoustic sequence was inverted and summed with the average detector voltage of the positive sequences. This produced a single dimension which was used to construct the BEST-SCARS hash table for classification of each sample group and identifying false samples. A hash table comprising 2000 distances was used for each group. In total the final hash table was an 18000 x 3 matrix. A hash table was constructed for each of the 1, 10, and 100 frequency acoustic sequence models. False samples were randomly scanned with the training set

samples. Estimated BEST MSDs were calculated for each voltage spectrum using the hash table approach.

Results and Discussion

FTARS vs. SCARS. The FTAR spectra were smoothed with a cubic spline operation. The mean smoothed spectra for each pure component over a frequency range of 3.1 to 4.4 kHz are depicted in Figure 2.3. The large resonance peaks and the clear differences between the pure component spectra made this a useful region for classification studies. Spectra over the selected frequency range were investigated by PCA to determine the relationships between the samples. The first three PCs accounted for 91.2% of the variation in the sample set (see Figure 2.4). It is evident from the figure that separation between sample groups was achieved. The loadings corresponding to the first three PCs were used to construct acoustic sequences for the SCARS method. Three different frequency models were investigated comprising six acoustic sequences each. The summed voltage signal corresponding to a negative loading vector was subtracted from the signal corresponding to the positive loading vector. This resulted in a three dimensional data set analogous to PCs and was used for classification. Figure 2.5 illustrates the separation of the voltages acquired from the single frequency acoustic sequence. Comparison of Figures 2.4 and 2.5 indicate that the separation achieved by performing PCs on the full data set can be accomplished at the detector by weighting the excitation signal. A leave-one-out cross validated BEST method was used to quantify the differences between FTARS and the SCARS methods. The performance of the classification can be measured by calculation of sensitivity and specificity according to:

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \quad (2.7)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \quad (2.8)$$

Where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. Table 2.2 indicates the performance statistics for classification with PCA and SCARS using 1, 10, and 100 factors. All samples were correctly classified in all cases resulting in a sensitivity of 100%. However, specificity decreases somewhat for the SCARS model. The decrease in specificity for the SCARS method is a result of a decrease in inter-cluster distances between some groups. For example the proximity of sample 5 (50:50 ratio of tag to RSV) to its pure component samples (1 and 2) decreases in the SCARS model. This in turn increases the chances of calculating a false positive between these groups. The high sensitivity of the SCARS model, which is impart due to the use of single samples for each group, suggests that it may be possible to decrease the MSD cutoff for some groups in order to achieve better specificity. Figure 2.6 illustrates the changes in mean intra- and inter-cluster distances between each model. For SCARS factor models the overall mean inter-cluster distances increase while the mean intra-cluster distances decrease. The largest mean inter-cluster distance is found in the largest single factor loading model. This is because the loadings in PCA are calculated such that their magnitude corresponds to the magnitude of the largest variation at each frequency. In Figure 2.3 it is clear that the magnitudes of the differences between the spectra are different at different frequencies. Therefore the PC loadings are also different in magnitude at different frequencies. A single loading will maximize the variation of the sample set at a single frequency. It is not always the case that there is significant

variation between all sample groups at a single frequency. Where variation between samples is not maximized it becomes more difficult to distinguish between them at a single frequency. The result is that some spectral groups are maximized and some are minimized. In Figure 2.6 the maximization of the MSDs between some of the groups clearly outweighs the minimizing effect for the single factor model. While there is little difference between the 1, 10, and 100 factor loading models, there appears to be a trend in the decrease of inter-cluster distances and increase of intra-cluster distances. This trend can be attributed to weaker signals being overwhelmed by noise. When 100 factor frequencies are used the amplitude of the frequencies corresponding to a portion of the factors can be significantly weak in comparison to the top 1 or 10.

BEST-SCARS for False Sample Identification. The BEST-SCARS method operates on the assumption that samples of the same composition will cluster together in hyper space. The precomputed hash table distances are calculated from a training set that should contain as much of the variation of the sample set as possible. This approach effectively draws a hypothetical line around each sample cluster in hyperspace. When a new sample is scanned it is projected into hyperspace; samples that fall outside of the line are identified as false samples. So long as the proportion and number of frequencies used in the SCARS method can distinguish between substances, any sample not included in the training set will be identified.

To determine the ability of the BEST-SCARS method for identifying false samples, three additional samples outside of the sample set were included. Figure 2.7 indicates the separation of the false samples achieved with the single factor BEST-SCARS model. Sensitivity and specificity were used to quantify the performance of the

BEST-SCARS method for each factor model in identifying the false samples. Table 2.3 indicates the performance statistics for identifying false samples using the hash table approach. There is a trade-off between sensitivity and specificity for any measure. The optimum sensitivity would be 100% indicating that no false samples were ever classified as part of a training set. In each case sensitivity was greater than 91% and specificity was always 100%. The sensitivity increased for the 10 factor model because more frequencies were needed to separate the false samples from the training set samples. The decrease in sensitivity in the 100 factor model can again be contributed to weaker signals in the excitation.

Only one false sample was identified as belonging to a training set. Table 2.4 reports the average BEST-SCARS MSDs for the single frequency excitations. An MSD greater than 3 (99.8% confidence limit) indicates that samples are separable. The false sample comprising a third mixture of glucose, RSV and Stilbene (FS1) was misidentified as sample 7. Because the intra-cluster distances of the training set sample groups are low (average of 1.15) it is possible to decrease the MSD cutoff for identifying false samples. An MSD cutoff of 1.96 (95% confidence limit), which is smaller than the lowest inter-cluster distance (2.13), would increase the sensitivity for identifying false samples to 98.15%. This would decrease the specificity (96.18%) but would reduce the chances of producing contaminated substances which could lead to costly recalls.

Speed of method. The BEST-SCARS method is much faster on a digital computer than analysis of ARS data with the full BEST algorithm for identification of false samples. In BEST-SCARS, MSDs are precomputed and each new sample scanned is compared to a table of values. The BEST-SCARS method compared each sample

scanned to a hash table of 2000 distances for each group. In total 18000 distances were compared, and the closest distance in the direction to the sample was chosen to estimate the MSDs of the samples in Table 2.4. Identification of all false samples was completed in an average of 0.135 s. Performing the full BEST algorithm on the same sample set required an average of 5.515 s.

Conclusion

Bootstrap error-adjusted single-sample technique sequencing computational acoustic resonance spectroscopy has been demonstrated to perform comparably with Fourier transform acoustic resonance spectroscopy. ARS is nondestructive, rapid, and inexpensive. Coupled with the ability of the BEST-SCARS method to identify false samples rapidly, it is a useful choice for PAT applications. Operating as a PAT, a BEST-SCARS sensor could monitor several ingredients on a manufacturing line. If a foreign ingredient is added or a substance is combined at the wrong time, a false sample situation would arise and the sensor system could ask for operator assistance. The SCARS acoustic sequences could be tailored for various common drug compositions. Tailored excitations for common pharmaceutical ingredients could be stored as MP3s and downloaded with their classification hash table. If a company wishes to make aspirin they would simply download the appropriate excitations and hash tables needed to validate their product.

Chapter Two Tables

Sample	D-Tag	RSV	Stilbene
1	100	0	0
2	0	100	0
3	0	0	100
4	50	0	50
5	50	50	0
6	12.5	12.5	75
7	75	12.5	12.5
8	12.5	75	12.5
9	33.3	33.3	33.3

Table 2.2: Performance statistics for BEST classification of PCA data and SCARS voltage data from sequences constructed from frequencies corresponding to the top 1, 10 and 100 loading factors.

Model	PCA	1 Factor	10 Factors	100 Factors
Sensitivity (%)	100	100	100	100
Specificity (%)	95.83	93.40	92.60	94.79

Table 2.3: Performance statistics for BEST-SCARS identification of false samples from sequences constructed from frequencies corresponding to the top 1, 10 and 100 loading factors.

Model	1 Factor	10 Factors	100 Factors
Sensitivity	92.60	94.44	91.67
Specificity	1	1	1

Table 2.4: Average BEST-SCARS calculated MSDs for false samples to each training set group for the single factor acoustic sequence. Sample clusters greater than 3 MSDs are considered to be from different groups (99.8% confidence limit). FS1 is a third mixture of D-glucose, RSV and Stilbene, FS2 is pure D-glucose and FS3 is vitamin B-12.

Group	1	2	3	4	5	6	7	8	9
FS1	4.61	22.82	28.28	7.99	5.76	14.65	2.13	8.65	3.59
FS2	4.85	21.83	38.29	12.85	5.69	23.01	3.72	9.25	10.36
FS3	7.42	29.16	27.89	6.38	8.24	13.59	3.26	12.38	11.44

Chapter Two Figures

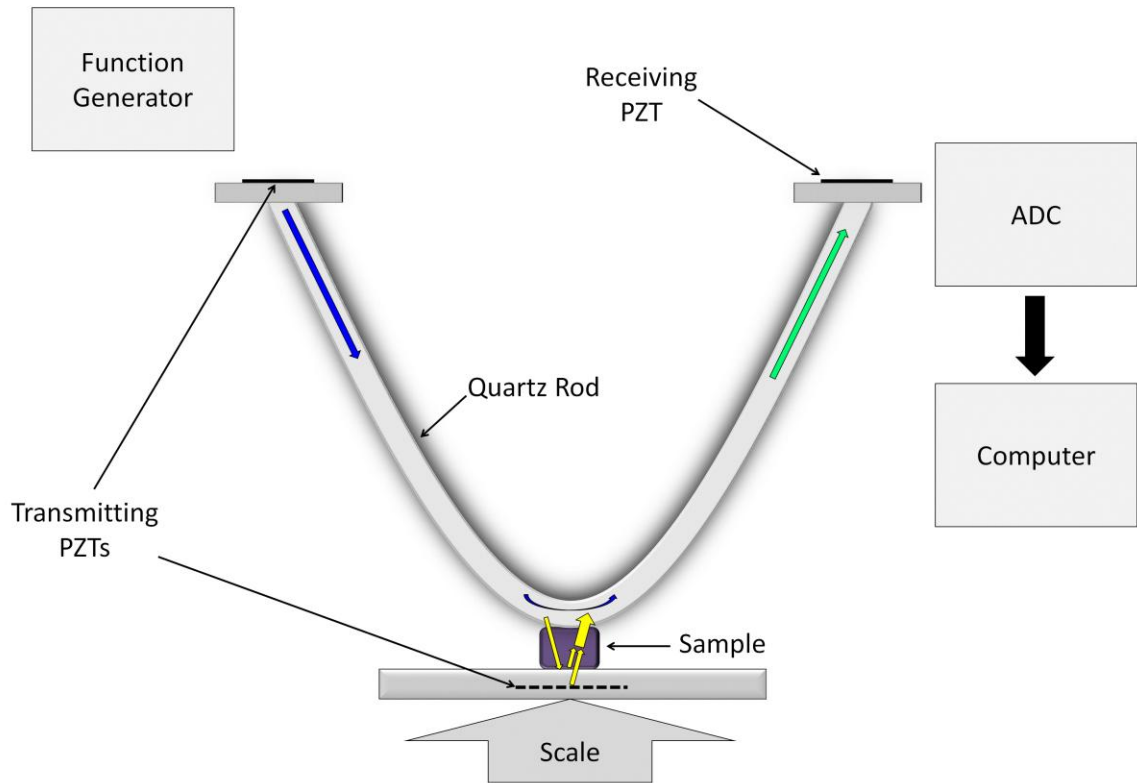


Figure 2.1: Schematic representation of the ARS. A function generator is used to provide a broadband signal to both transmitting PZTs. The emitted signal interacts with both the quartz rod (blue arrow) and sample (yellow arrow). The second transmitting PZT acts to boost the resonance signal from the sample. Constructive and destructive interactions between the acoustic waves of sample and quartz rod produce the standing wave (green arrow) at the receiving PZT which is read by the ADC and stored in a computer for processing.

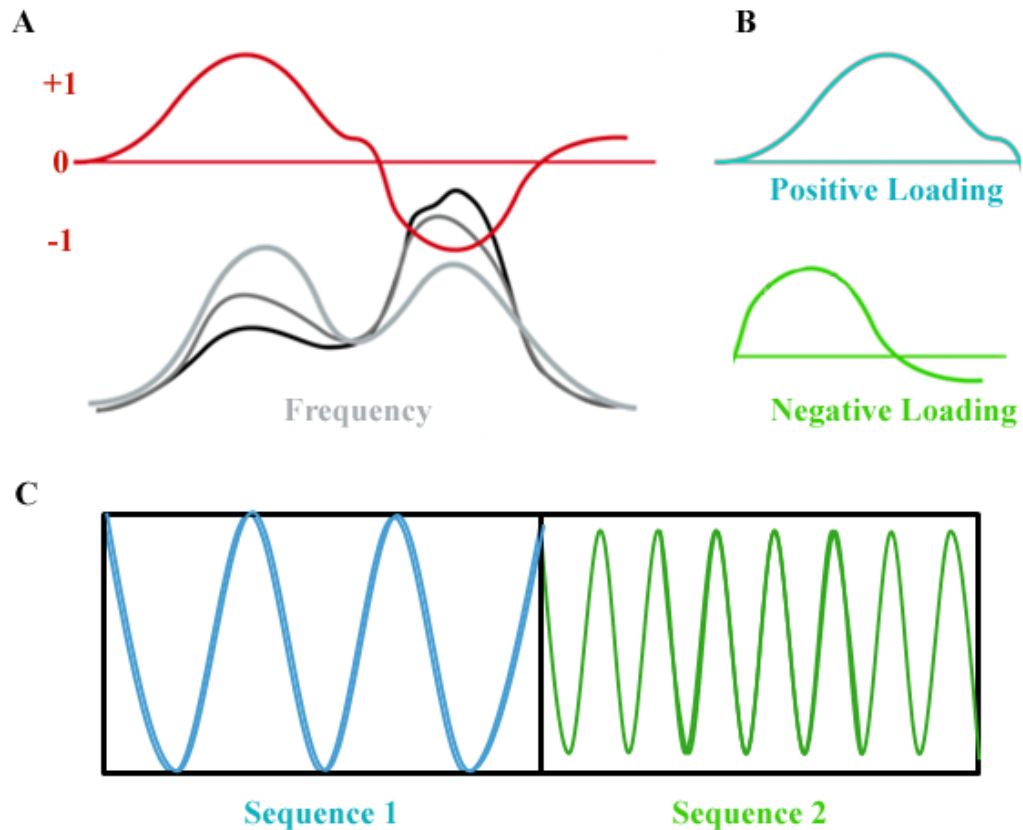


Figure 2.2: The frequency ranges comprising the largest spectral variation in a data set can be visualized in their PC loadings (A). Because loading values are weighted in both the positive and negative direction they must be separated and the negative values inverted (B). A single frequency pulse sequence is constructed from the frequency corresponding to the largest loading value. When multiple frequency components are used the amplitude of each frequency comprising the pulse sequence corresponds to the magnitude of the corresponding normalized loading value. For each PC loading two pulse sequences must be transmitted independently through the sample to prevent cancelation of the integrated signal (C).

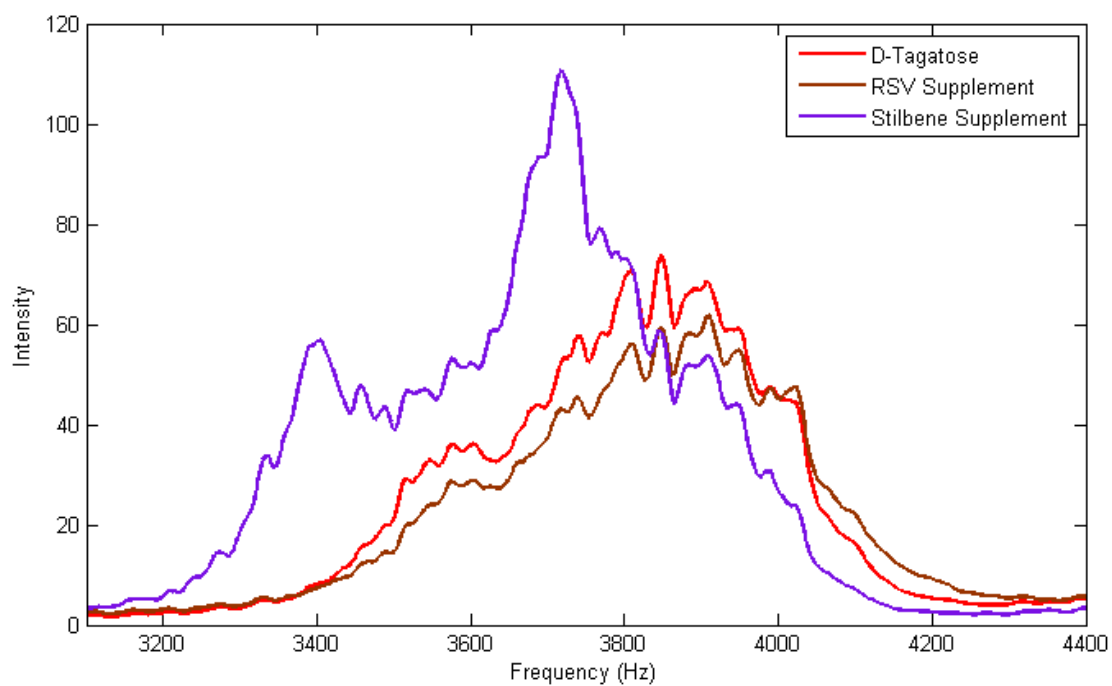


Figure 2.3: Pure component spectra over the frequency range of 3.1 – 4.4 kHz. A clear difference was observed between the pure component spectra. Each spectrum is the mean of 4 spectra taken from 12 scans of the same analyte.

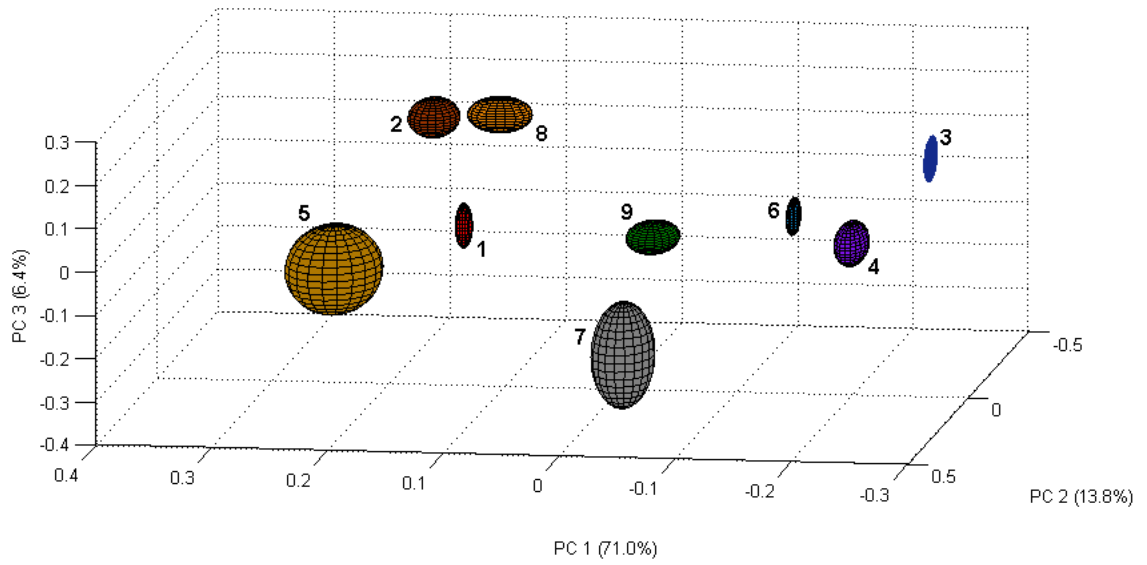


Figure 2.4: An XYZ scatter plot of the top three PCs. PCs 1-3 captured 91.2 % of the sample variation. The ellipses indicate a one standard deviation contour level.

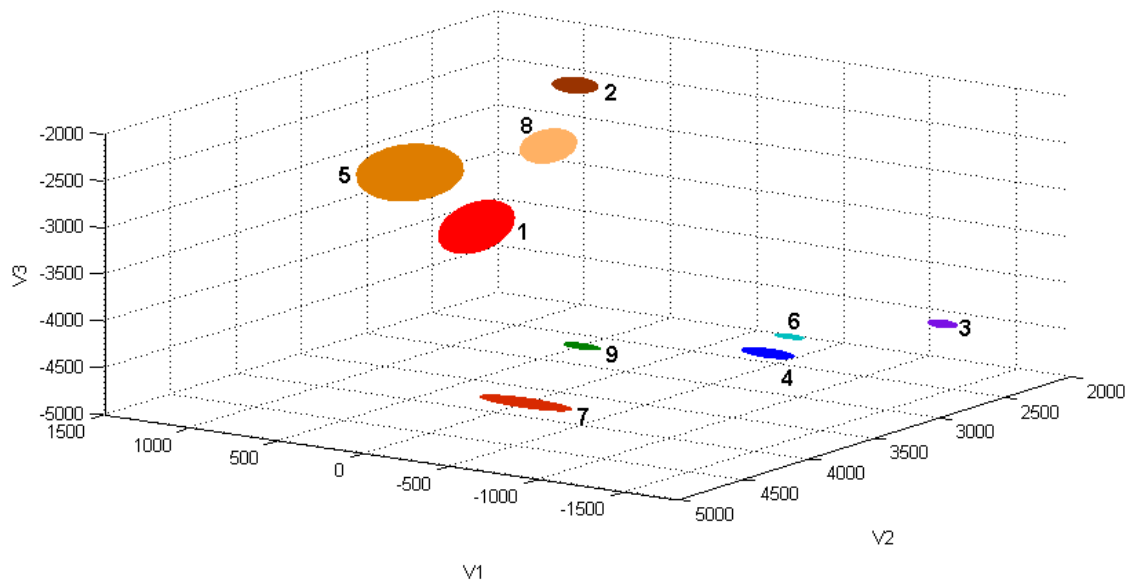


Figure 2.5: XYZ scatter plot of the voltage scores from the single frequency SCARS voltage data. The ellipses indicate a one standard deviation contour level.

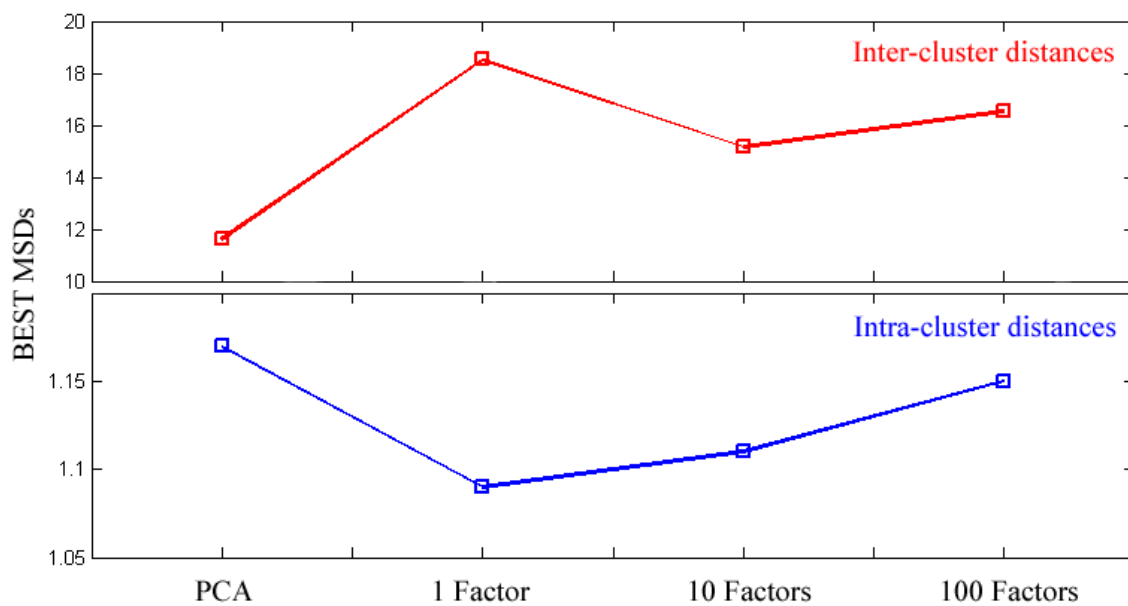


Figure 2.6: Mean BEST MSDs for inter- and intra-cluster distances calculated from the full BEST algorithm for the Full spectrum PCA data and from the BEST-SCARS on 1, 10, 100 and 1000 top factor loadings.

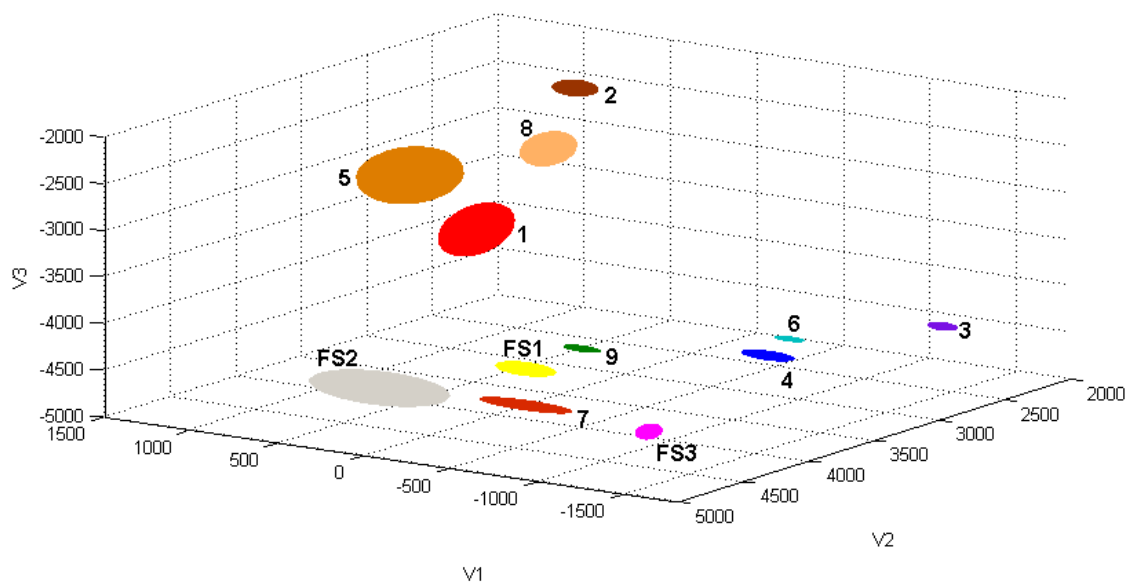


Figure 2.7: XYZ scatter plot of the voltage scores from the single frequency SCARS voltage data. The ellipses indicate a one standard deviation contour level. FS1 is a third mixture of D-glucose, RSV and Stilbene, FS2 is pure D-glucose and FS3 is vitamin B-12.

Section III: Remote Alcohol Sensing

The abuse of alcohol (ethanol) is a major problem in the United States for health and economic reasons. More than 125 million Americans aged 12 or older are alcohol consumers⁶⁸. Alcohol abuse can lead to heart and liver problems, increased risk-taking behavior, and reduction of the ability to operate a motor vehicle⁶⁸. According to the National Traffic Highway Safety Administration's Fatality Analysis Reporting System (FARS), in 2004, 39.5 percent of all auto crash fatalities were the result of alcohol impairment⁶⁹. The problems arising from the abuse of alcohol are not limited to the road and affect the well-being of many young Americans as well. Alcohol has been shown to lower inhibitions and increase risk-taking behavior⁶⁸. In 2001, over 1,700 alcohol-related unintentional deaths occurred among college students, and there were approximately 696,000 assaults by students who had been drinking⁷⁰. Economically, alcohol abuse harms entire communities, costing taxpayers huge amounts of money spent in order to protect them from those who are abusing alcohol. The total yearly economic cost of alcohol abuse in 1998 was estimated to be 184.6 billion dollars⁷¹.

Treatments for abusers of alcohol, including Alcoholics Anonymous (AA) and other behavior modification-style organizations, are easy to find. However, the bulk of alcohol-related problems are caused by individuals who are not seeking help. Approximately 39 percent of the 400,000 Americans admitted to alcohol treatment programs are there due to a court order and not by personal choice⁷². The need to remain alcohol-free in order to remain free from incarceration provides an incentive for individuals to lie about their consumption to counselors and parole officers. A similar problem arises for the scientific community studying the effects of alcohol treatment

programs. Evaluating treatments requires a continuous monitor of subjects' blood alcohol content (BAC), breath alcohol content (BrAC, known to be highly correlated to BAC⁷³), or actual consumption. Current alcohol content monitoring techniques include the sampling of breath, urine, saliva, or blood followed by a different analysis to assess the alcohol level in each. Blood testing is the most sensitive method, but the test can suffer from temporal distortion due to the processing time needed to obtain the result. Samples can ferment on their own over time, leading to inaccurate measurements of alcohol consumption. Temporal distortion is problematic in the research setting when a time-sensitive result is needed. Urine testing is the most inexpensive of the techniques, but it tests for alcohol in the system in the past five days, not in real-time⁷⁴. Techniques that integrate consumption over long periods of time complicate experiments designed to test the effects of treatment on acute alcohol abuse. Contamination, dilution, and tampering of the samples are other common problems in urine testing⁷⁴⁻⁷⁵. Breathalyzers are the most common on-site monitoring technique and are generally used by law enforcement on the public. The breathalyzer is the only noninvasive technique that can be used to monitor real-time alcohol levels, but it has many drawbacks. Some breathalyzer devices assume a hematocrit (cell volume of blood) of approximately 47 percent, when in reality it ranges from 37 to 52 percent⁷⁶. Testing an individual with a hematocrit level below the assumed value will result in a false positive reading on the breathalyzer⁷⁶. Breath testing can also lead to false positives when blood, alcohol, or vomit is present in an individual's mouth. One of the most significant drawbacks of the breathalyzer is that the individual must be a willing participant in the measurement process.

One approach to monitoring alcohol abuse is through the use of wristband fuel cells or implants. The use of the Secure Continuous Remote Alcohol Monitoring (SCRAM) system has become common for those convicted of DUI. The SCRAM device monitors transdermal alcohol content (TAC) in the body. The SCRAM is attached to the offender's ankle for typically a period of a month. However, problems have been noted with use of the SCRAM device. An offender placed in an alcohol supervision program that uses a SCRAM device may not use any products such as mouthwash or household cleaners that contain alcohols, because these products can cause the device to produce a false positive. The device monitors attempts to bypass it by monitoring body temperature. If a sudden temperature change is recorded at the detector, an alert will be sent to the monitor. A problem with all monitoring methods discussed so far is that the subject is aware that he or she is being monitored.

Chapter Three- Surreptitious Remote Sensing of Blood Alcohol Concentration: Integrated Near-Infrared Spectral Imaging and Laser Speech Detection

Introduction

A noninvasive and surreptitious measurement of BrAC is of interest because of the many problems regarding the existing real-time and more intrusive methods. If a subject is unaware of when the measurement is being made, it becomes more difficult to attempt to defeat the measurement. Figure 3.1 illustrates the use of MFC-NIR hyperspectral imaging and laser speech detection for unobtrusive monitoring of a subject. This paper describes experiments to estimate BrAC using NIR hyperspectral imaging based on molecular factor computing (a method of integrated sensing and processing) and speech analysis. These approaches have the potential to estimate alcohol impairment unobtrusively, remotely, and in real-time.

Near-Infrared (NIR) Imaging. NIR spectroscopy has been demonstrated to be a valuable analytical tool for the simultaneous determination of multiple chemical components in mixtures. As a result, NIR spectroscopy is used throughout the biotechnology and pharmaceutical industries⁷⁷⁻⁸². Pulse oximetry, a medical application of NIRS, is very common in hospitals⁸³. The NIR region of the electromagnetic spectrum offers advantages for use in biological systems. NIR radiation is able to penetrate through the dermal layers of skin and has been shown to measure accurately blood levels of analytes *in vivo*⁸⁴⁻⁸⁵. Thus, diffuse reflectance NIR imaging is a candidate for noninvasive determination of blood alcohol. The combination of NIR imaging with molecular factor computing (MFC) filters offers several advantages over the traditional

approach to NIR imaging⁸⁶. Advantages of using an MFC filter system with NIR imaging include a decrease in computational demand by integrating the sensing and processing directly into the transducer (the integrated sensing and processing, or ISP, advantage), the Fellgett advantage (acquisition of all wavelengths simultaneously permits higher signal-to-noise ratio for a given acquisition time), and the Jacquinot advantage (the absence of monochromator slits allows higher optical throughput). The use of a small number of MFC filters also allows for a more robust instrument at a lower cost. The concept of molecular factor computing has been demonstrated in previous work⁶⁶.

NIR hyperspectral imaging creates large volumes of data because images are acquired over many wavelengths where the analyte is expected to have signals. Most images also contain a substantial number of pixels, and many images are often collected to form a video stream. The data sets generated by hyperspectral imaging can be difficult to analyze. To this end, chemometric methods such as factor analysis are commonly used to reduce large data sets into factors and corresponding weights that relate to the observed variance in the data set. MFC-NIR hyperspectral imaging exploits these chemometric methods and can be an effective alternative to conventional hyperspectral imaging^{79, 87-88}. MFC essentially incorporates chemometrics directly into the physical design of the wavelength selector, thus integrating the sensing and processing at the detector itself^{66, 63}. MFC filters are selected to have transmission spectra that match the weights generated from factor analysis of a calibration set. Thus, the light passing through a MFC filter is effectively “weighted”, and the corresponding signal at the detector is proportional to a factor. The use of weighted filters allows for factor analysis to be performed at the speed

of light. The chemometric method we have used to perform MFC is principal component analysis (PCA).

Molecular factors are computed by first generating a training set of spectra over a concentration range of interest. An orthogonal linear transformation (PCA) of the raw spectra is done by a singular value decomposition of \mathbf{A} according to Equation (3.1):

$$\mathbf{A} = \mathbf{S}\mathbf{L} \quad (3.1)$$

where \mathbf{A} is an $m \times n$ matrix of original spectra, \mathbf{S} is an $m \times n$ matrix of the component scores (PCs), and \mathbf{L} is a $p \times n$ matrix of loadings (weights). The dimensions of the matrices are dependent on the number of samples, m , and the number of predictors, n , and p is the number of PCs. The decomposition of \mathbf{A} is calculated such that the total variation of the training set is found in the component scores (PCs). The first PC contains the highest amount of variation, the second contains the next highest, and so on until the total variation is explained. The regression of \mathbf{S} , as shown in Eq. 3.2, produces a robust model for predicting the concentration \mathbf{C} in a complex sample, where \mathbf{R} is an $m \times p$ matrix vector of regression coefficients, and \mathbf{D} is the residual.

$$\mathbf{C} = \mathbf{R}\mathbf{S} + \mathbf{D} \quad (3.2)$$

It is expected that a pure two-component system can be represented by one PC. However, due to instrumental noise, chemical inconsistencies, and spectral overlap, more than one PC may be needed. MF filters are selected that have transmission spectra which closely match the loadings that correspond to the PCs used in the training set validation. Two filters are required for each PC loading because a MF filter cannot be negative. The product of the loading and spectral amplitude summed over all wavelengths produces a voltage at the detector according to Eq. (3.3).

$$V_i = \sum_j^J L_j A_{ij}$$

for $j = 1, 2, \dots, n$

$i = 1, 2, \dots, m$ (3.3)

where \mathbf{L} is the loading at wavelength j , \mathbf{A}_{ij} is the spectral amplitude of the i th mixture at wavelength j , \mathbf{V}_i is the detector voltage, and j is an index over J wavelengths. All wavelengths passing through a MF filter are weighted according to \mathbf{L} . The resulting \mathbf{V} is directly proportional to the corresponding factor of \mathbf{L} and proportional to the sample properties (BrAC) being investigated. Figure 3.2 illustrates the use of MFC filters. A theoretically perfect MFC filter would have a transmission spectrum that perfectly matches a positive or negative loading profile of the sample. In most cases, a perfect MFC filter cannot be constructed. In cases where a perfect MFC filter cannot be constructed, partially modeling a PC loading may still give acceptable results. In previous work, liquid filters were used to determine ethanol and water concentrations⁶⁶. In this work, solid polymer filters were used that were composed of non-alcoholic functional groups.

Speech Detection. Determination of BrAC using speech is based on changes in phonemes that occur after consumption of alcohol. It has been shown that an individual's speaking fundamental frequency changes with alcohol consumption⁸⁹. Language comprises various sounds—the smallest units that define spoken word are known as phonemes⁹⁰. In the United States, the English language consists of approximately 44 phonemes⁹⁰. While the number of phonemes in other languages is different from English,

some speech changes would still be expected to be associated with alcohol consumption. Speech can be recorded surreptitiously and remotely, and there are various instruments made for this purpose, such as parabolic microphones (Dan Gibson Parabolic Microphone, Electromax International, Inc. Houston, Texas) used in television and broadcasting. However, there are other less obtrusive methods available for monitoring speech. A laser microphone-listening device outside of a building, such as the one offered by Brinkhouse Security (New York, NY), could be used to remotely monitor conversations inside a building, as long as glass windows are present in the building walls. Using a laser microphone method, law enforcement personnel could monitor conversation inside a restaurant, bar, or car, making determinations of which patrons are over the legal limit for BAC. As the patrons went to their car and left, they could be stopped by police.

The following work describes preliminary research to determine which remote-sensing modalities can provide sensitive and specific detection of BrAC.

Materials and Methods

Patient Screening. The unobtrusive remote-sensing of human breath alcohol content study was conducted under IRB-approved (approval number 07-0417-F1V) protocols. Many individuals were screened for the study. However, due to inclusion criteria, only 5 subjects were allowed to participate. At screening potential subjects were asked to provide a complete medical history and were questioned about the use of concomitant medications. Inclusion criteria were male or female subjects between the ages of 21 and 45 who were not alcohol naïve, testing zero BrAC at the screening prior to the

commencement of the experiment, consuming no more than one standard drink of alcohol a day over the past 30 days, passing a physical examination and laboratory assessment at the initial screening, and having a body mass index (BMI) between 20 – 30 and total body mass no more than 100 kg. Criteria for exclusion from the study included subjects under the age of 21, consuming two or more standard drinks of alcohol a day over the past 30 days or have never consumed alcohol, recovering addicts, pregnant or breastfeeding women, liver disease or impairment, diagnosis of diabetes, present at screening with black or blue marks on their skin, present at screening with evidence of intoxication, withdrawal, confusion, severe depression or anxiety, immediate family members of subjects with a history of alcoholism or alcohol abuse, using other drugs that interact with alcohol, subject history of or current abuse of illicit drug or prescription drugs, smoking, or subject enrollment in another investigational device or investigational new drug study that had not been completed by the subjects' required follow-up period. Subjects were asked to complete the Alcohol Use Disorders Identification Test (AUDIT) and to fast the night before testing⁹³.

Clinical Trial Procedure. Subjects were administered pharmaceutical grade ethyl alcohol (195 proof) diluted in lemon soda according to total body mass based on the following dosing equations⁹⁴:

$$\text{_kg} \times 0.83 \text{ ml/kg} = \text{_ml of alcohol} \quad (3.4)$$

$$\text{_ml of alcohol} \times 3(3 \text{ parts lemon soda}) = \text{ml of lemon soda} \quad (3.5)$$

Generally, this resulted in a dose of 50 to 60 mL of ethyl alcohol to the patients. Subjects arrived the morning after fasting overnight and were given a breathalyzer test to insure an

initial reading of 0.00 % BrAC. Reference measurements were made at this time with the remote speech and NIR MFC devices. The subjects were given 20 minutes to consume the alcohol mixture at which time the patients rinsed their mouth twice with water (required for the initial measurement of BrAC). BrAC measurements were taken 30 minutes after administration (20 minute consumption period + 10 minute delay) of the alcohol dose. Remote speech and NIR MFC video measurements followed and were acquired at 10-minute intervals. Standard BrAC reference measurements were taken at the beginning of each 10-minute interval using a Breathalyzer (Draeger AlcoTest 7410 Plus). The test subjects agreed not to leave the testing room until they scored two consecutive BrAC measurements of zero. The protocol design and consent stipulations have followed all IRB regulations, guidelines, and local policies.

NIRS MFC Hyperspectral Imaging. An IRC-160 InSb focal plane array video camera (Cincinnati Electronics, Mason, OH) with molecular factor component (MFC) filters was used for imaging of the patients. The camera integration time was 12.96 ms, and the photon energy response was 1800-10,000 cm^{-1} . A rotating disk was fabricated to allow the different MFC filters to be rotated in front of the camera lens. Six filter materials were used for the study: polyvinyl chloride (PVC, Unbranded, McMaster-Carr), polycarbonate (PC, Lexan, Plaskolite, Inc.), Acrylic (AC, Unbranded), polymethylmethacrylate (PMA, Optix, Plaskolite, Inc.), combined gel filters CC20B and CC40G (BG, Kodak), and Gel A2 Pale Yellow (A2, Kodak). The light source was two 250W PC37771 lamps (General Electric, Cleveland, OH). Chemicals for the MFC filters were chosen using a genetic algorithm library search. In previous work, liquid filters were selected with the algorithm from a library of NIR transmission spectra containing 1,923

compounds (Wiley)⁶⁶. The genetic algorithm simply searched the database for up to ten compounds that provided a combined transmission spectrum that closely resembled a factor loading. The spectral library available for this research was composed of liquids, which would be harder than polymer filters to deploy in a clinical trial. For this reason, a smaller list of available polymers was ultimately used to create the MFC filters. Selecting from a smaller database of available filters limits the performance of the MFC hyperspectral imager somewhat. However, solid polymer filters make the system more rugged and easier to use in the field.

Subjects were never in exactly the same position for measurements. To compensate for changes in image contrast and intensity resulting from subject position changes and variable illumination over the acquisition period, pictures of two spherical silicon dioxide reflectance standards⁷⁷ (one high reflectance and one low reflectance) were captured in each image. Images in the video stream were made comparable by multiplicative scatter correction^{92, 87} based on the reflectance standards, so that the standards appeared identical throughout all images. A flat image reference standard (Kodak Gray Card, Rochester, NY) was also used to correct for temporal and spatial inconsistencies in detector response across the pixels of the detector array. Two light sources were employed during image acquisition to reduce shadowing and increase the signal-to-noise ratio of the data. The lights were placed on each side of the camera and approximately 135 degrees relative to an orthogonal line connecting the patient and camera lens. Images were obtained with each MFC filter with the light sources turned off and with them on (12 total spectral images) to correct for ambient lighting from the room

and blackbody radiation from the patient. The entire video acquisition process took only 2 minutes.

Laser Speech Instrumentation. The remote speech detection instrumentation was fabricated in-house and was based on Michelson interferometry. The interferometer utilized a battery-powered laser pointer (635 nm) as the source to simplify aiming. (Production versions will use an invisible laser wavelength within the range of the NIR camera to enable aiming.) The interference fringe pattern was detected with a phototransistor and amplifier using a soundcard (M-Audio, Avid Technology Inc) and was recorded into Cool Edit Pro (Syntrillium Software Corp.) at a sample rate of 44.1 kHz. The laser beam was aimed at a glass target positioned 2 to 3 feet from each patient. The sound from patients' speech induced vibrations in the glass causing constructive and destructive interference patterns at the phototransistor, which were amplified and recorded. The changes in the interference patterns were stronger at frequencies where the glass target had resonances. A unidirectional microphone (ECM-330, Sony) that lacked the resonances of the glass was also used as a reference and placed next to the reflecting glass. This reference was used ad hoc to estimate the resonances of the remote glass. The test subjects were given a list of 13 words to read (table, flat, feet, pet, light, bit, bone, hot, future, thumb, boot, soil, saw) containing common phonemes in the English language.

Data Analysis. Analytical software was written in MatLab 7.1 (The Mathworks, Inc.). Principal component analysis (PCA)¹⁵ was used to analyze all data except for MFC

voltage data. The multivariate regression method was cross-validated in predicting BrAC using both NIR MFC images and speech data. The F test and the standard error of performance (SEP) were calculated from the validation samples⁹⁵⁻⁹⁶.

The NIRS video consisted of indexed color images with a resolution of 72 pixels/inch and dimensions of 720x480 pixels (width by height). The images were imported into MatLab 7.1 where they created a matrix of intensity levels with a size of 720x480x3 (rows by columns by dimensions). The dimensions corresponded to the red, green, and blue (RGB) color space where the color intensities for a specific pixel were combined to give 1 of 256 possible colors⁹⁷. The subject's face fit the vertical length of the image and was constant. However, horizontal movement of the subject's face throughout the testing period could not be avoided. Therefore it was necessary to crop portions of the subject's face because of movement, dead areas in the frame, and areas where the signal-to-noise ratio was very low due to highly reflecting materials such as hair. Each subject's face was cropped in four areas: forehead, under the eyes, and lower face as depicted in figure 3. In some cases the use of lower face data was prevented by facial hair, and forehead data was limited where the subject had low hanging hair. The cropped images were separated into three layers representing RGB, then gray and multiplicative scatter correction was performed. To determine if the voltage corresponded to a PC score, each color layer of the cropped images was averaged to produce a single voltage score. Multivariate regression was performed on the voltage data from each cropped facial region. All possible combinations of three, four, and five MFC filters of the total six were used to determine if a correlation existed. Because the MFC filters were not perfectly weighted functions of the ethanol in humans, PCA and

multivariate regression was also performed on the filter data. Leave-one-out cross-validation was used to determine the prediction ability of the model for BrAC values from NIRS imaging.

Cool Edit Pro was used to apply noise reduction to the microphone speech data only. This was done by creating a noise profile for each patient. The noise profile was created by sampling the background noise in the audio track (i.e. before the subject began speaking). The background noise sample was applied to the full waveform with the noise reduction option using a FFT size of 12000 and a precision factor of 10^{-98} . The data were imported into MatLab following noise reduction. The speech data was analyzed in two ways: analysis was completed on full phrases (i.e. all words together in sequence) and on individual words. To accomplish this, the time domain data was cropped so that there was one data set for all combined words and one data set for each of the 13 words spoken by each subject. All time domain data was Fourier transformed into the frequency domain. PCA was applied to all data sets and multivariate analysis was used to determine if correlations between speech and BrAC existed. All multivariate models were validated using leave-one-out cross-validation.

Results and Discussion

NIR Image Data. The purpose of the MFC filter is to increase S/N and reduce the numeric processing associated with an analytical method. In theory, if the MFC filters had been perfectly weighted functions of the ethanol in humans, the resulting intensity signals at each pixel of the images would be proportional to ethanol concentration in blood and no other calculations would need to be made. In this case there were six MFC

filters that yielded six pseudo PC scores (voltage scores). Calculation of the voltage scores from the MFC filters required sampling several pixels at different facial regions, as described previously. The pixels were averaged together for each image and used as the single voltage score for the corresponding filter. This produced one data point per image per filter. In the end there were six voltage scores for each sample concentration. However, this analysis did not produce highly correlated results probably due to the imperfect nature of the filters as factor loadings and the introduction of noise from patient movement. For example, as described in figure 3.2, two filters are required for each PC loading used. The MFC filters used in this experiment represented only the positive loadings. Figure 3.4 indicates the r^2 values from each subject using multivariate regression and leave-one-out cross validation for MFC image data. The standard error of prediction (SEP) for the correlations in figure 4 can be found in figure 3.5. Information on the facial regions used in the calibrations reported in figures 3.4 and 3.5 can be found in the technical report from reference 99.

Principal components and multivariate regression was performed combining all combinations of colors (RGB) and MFC filters together to find the best correlation. The regression correlations and cross-validation statistics for each subject from a full facial profile that mathematically describe the fitness of the model and the prediction ability of the model for BrAC values from PCA of MFC NIR imaging is indicated in figures 3.4 and 3.5. Post-processing factor scores (i.e., 6 scores instead of ca. 1000 wavelengths) by PCA gave the best results. Information on the color space and MFC filters in the PCA regression model can be found in reference 99. Image data were assembled together from all patients to make a pooled global calibration for testing the prediction ability

when the subject is unknown (i.e., not previously calibrated). Figure 3.6 A. indicates the global model predicted BrAC with $r^2 = 0.987$ and SEP = 0.0081 % BrAC.

Speech Spectral Data. The time domain data was Fourier transformed (FT) into a frequency domain. The FT was performed such that every frequency represented a data point (sample rate = 44.1 kHz, therefore each sample was a 1x22050 matrix). All spaces in the time domain data not containing speech were removed by editing so that the FT was a combination of all frequencies. Due to the large volume of data, PCs were calculated over a limited frequency interval where speech was expected to be found. Frequency differences were observed among the different phonemes associated with the words used in this study. Figure 3.7 illustrates the speech spectra from patient 5 obtained from the interferometer and microphone for words “boot” and “light”. The interferometer produced lower frequencies similarly to that of the microphone. However, it was found that the microphone captured higher frequency signals with a much better S/N ratio. The laser interferometer produced speech spectra that showed BrAC correlations with frequencies lower than those observed in the microphone spectra. This was likely attributed to resonance with the reflecting glass. Due to the nature of the recording device and room noise, the signal to noise (S/N) ratio of the laser interferometer was low. As a result of the laser interferometer detecting lower frequencies than the microphone, correlations with BrAC were found at much lower frequencies. The average frequency range associated with BrAC correlations to interferometer-recorded speech was 150 to 900Hz. The laser interferometer-recorded speech spectra produced individual correlations with $r^2 > 0.91$ and SEP < 0.017 % BAC.

Figures 3.8 and 3.9 depict the r-square and SEP statistics respectively, from multivariate regression and leave-one-out cross-validation observed for each word. An $r^2 = 0.806$ and $SEP = 0.017\%$ BAC was found for the average pooled interferometer spectral data. The relatively poor performance of the global model is most likely due to the small sample size ($n=5$) and may be attributed to the differences in individual vocal clarity and ability. Hollien et al noted that differences in fundamental frequency are found between men and women⁸⁹.

The microphone-recorded speech spectra produced individual correlations to BrAC of $r^2 > 0.95$ and SEPs $< 0.015\%$ BrAC. Figures 3.10 and 3.11 illustrate the microphone-recorded r-square and SEP statistics respectively, from multivariate regression and leave-one-out cross-validation. The average frequency range associated with BrAC correlations to microphone-recorded speech was 150 to 1050 Hz. A global model was calculated by pooling the microphone speech spectra for each word. An $r^2 = 0.805$ and $SEP = 0.016\%$ BrAC was found for the average pooled microphone spectral data. Figure 3.6 B and C illustrates the prediction capability for the global speech model for spoken words bit (interferometer recorded) and boot (microphone recorded), respectively. Again, the relatively poor performance of the global model is most likely due to the small sample size ($n=5$) and because of differences in vocal characteristics of each patient.

Conclusion

MFC-NIR hyperspectral imaging and speech are shown to correlate with BrAC. MFC-NIRS imaging returned a global correlation much higher ($r^2 = 0.987$, $SEP = 0.0081\%$)

than that of the speech detection method. Although the predictions made from principal component regression of the MFC filter data were somewhat better than predictions made from the voltage data alone, the results suggest that MFC may be capable of producing a more robust prediction. The MFC filters used in this research were selected from a small database of available materials. The number of filter materials available for MFC is high and better methods of filter selection may be needed for MFC to be a more useful technique. There may also be a market for production of solid state filters for MFC-NIR hyperspectral imaging. The advantages of MFC-NIR hyperspectral imaging over traditional NIR hyperspectral imaging include a decrease in the computational demand, shorter acquisition and analysis times, and higher signal-to-noise ratio. MFC also has the advantage of increased optical throughput because many wavelengths are acquired at the same time.

Problems with determining the portions of the patient's face to analyze increased measurement noise lead to a less accurate prediction model. This problem may be fixed by incorporating an algorithm to distinguish between exposed skin and hair (hair is highly reflective in the NIR).

Speech results for individual calibrations were shown to correlate well with BrAC. Although laser microphone-listening devices can be expensive, but we have shown it is possible to build a device for much less (under \$100). The global model for speech determination of BrAC is more difficult to find. However, the results of this study suggest that individual calibrations of speech can allow for its use in commercial industries.

MFC-NIR hyperspectral imaging and laser speech detection may be of use to law enforcement for noninvasive alcohol monitoring of subjects. Although, due to the ethical and constitutional dilemmas posed by this research, noninvasive alcohol monitoring may find better use in the pharmacology and alcohol treatment industries.

Chapter Three Figures



Figure 3.1: Remote alcohol sensing: a hypothetical situation where NIR laser beam is used to monitor speech from the glass on the table. MFC-NIR hyperspectral imaging is used for laser positioning and to determine the subject's BAC.

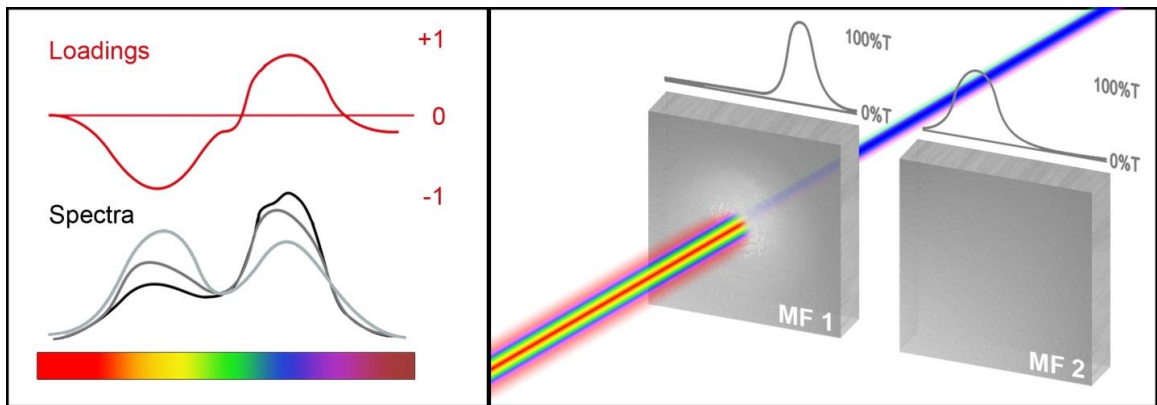


Figure 3.2: A set of training spectra is collected over many wavelengths. Loadings correspond to spectral regions where variation in analyte concentration is greatest. Molecular factor (MF) filters are selected to closely match a loading profile. Because MF filters cannot have negative responses, two MF filters are required. A broadband source is used to illuminate a target and the reflected light is passed through the MF filter. Only wavelengths corresponding to regions where spectral variability is greatest can pass through the filter. Because the MF filter is chemometrically weighted, the voltage response at the detector is proportional to the analyte sought.

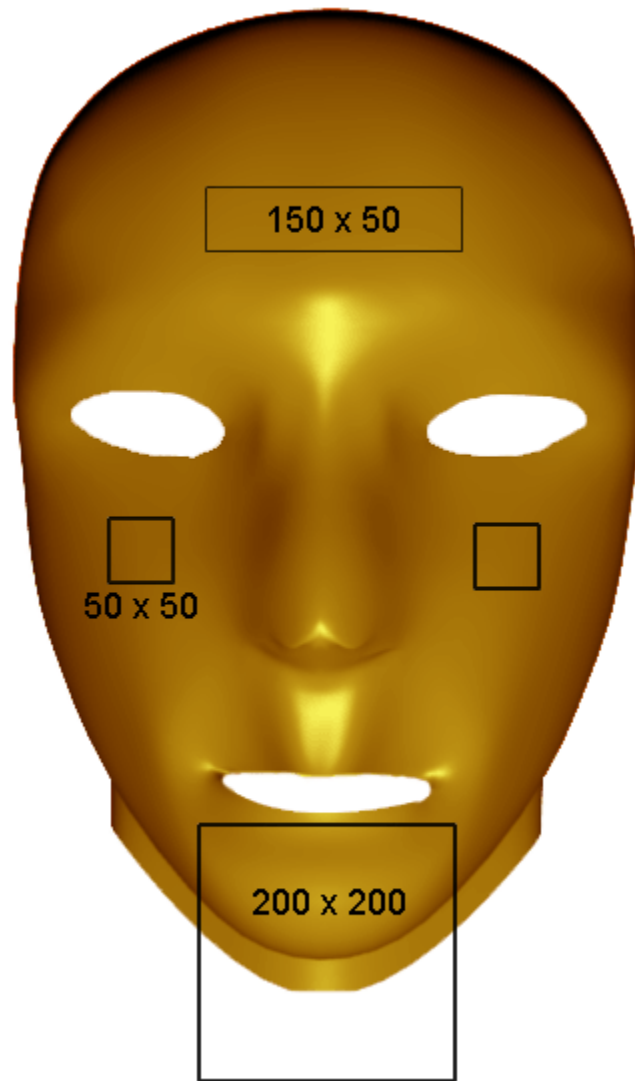


Figure 3.3: Due to movement of the subject over the acquisition period, each subject image was cropped in four areas: forehead, under the eyes, and lower face with horizontal and vertical areas indicated. The lower face region also included a portion of the subject's neck.

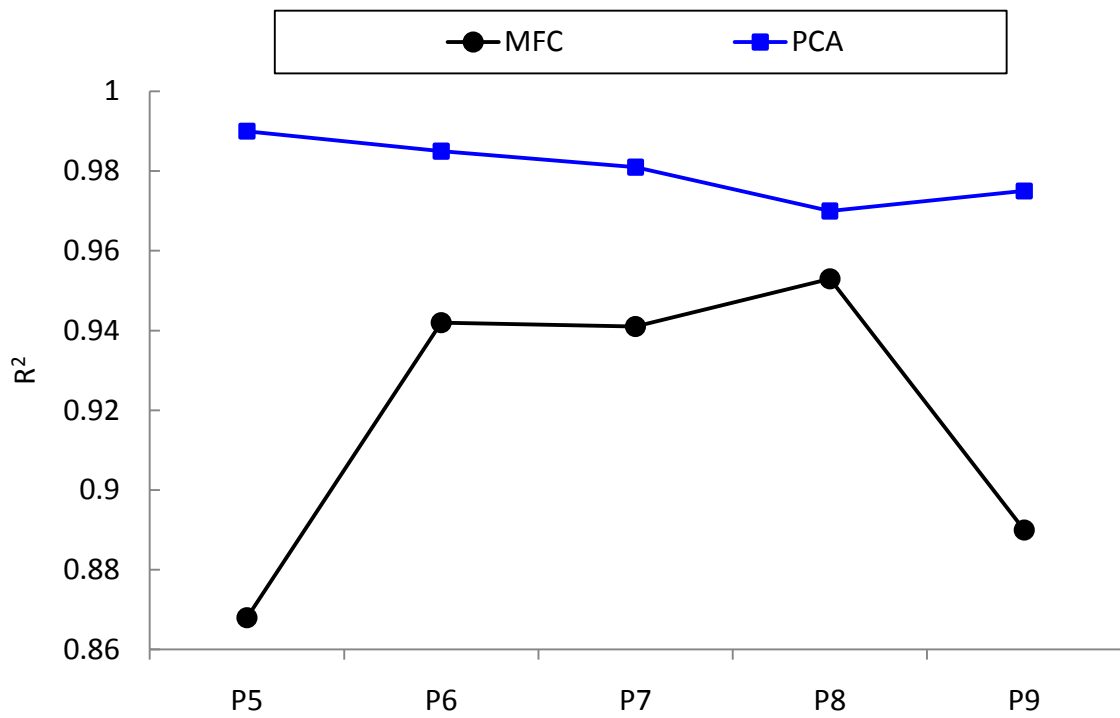


Figure 3.4: Comparison of MFC and PCA image calibration results from leave-one-out cross-validation of actual versus predicted % BrAC. P5 – P9 indicate patient ID.

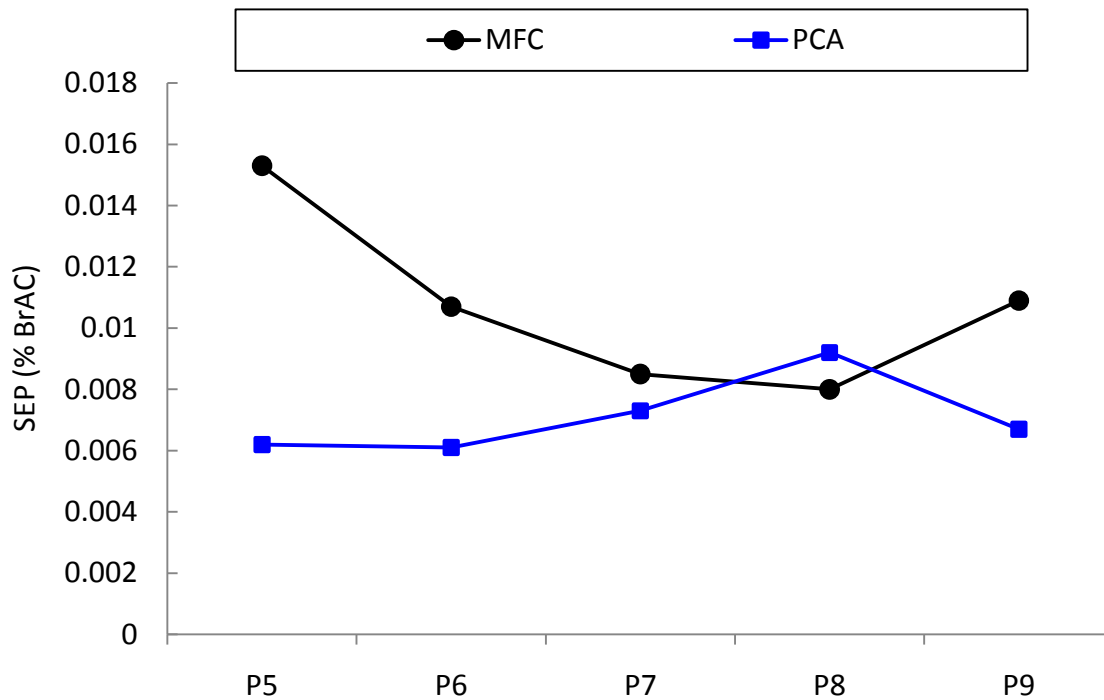


Figure 3.5: Comparison of PCA and MFC image calibration SEP values for leave-one-out cross-validation of actual versus predicted % BrAC. P5 – P9 indicate patient ID.

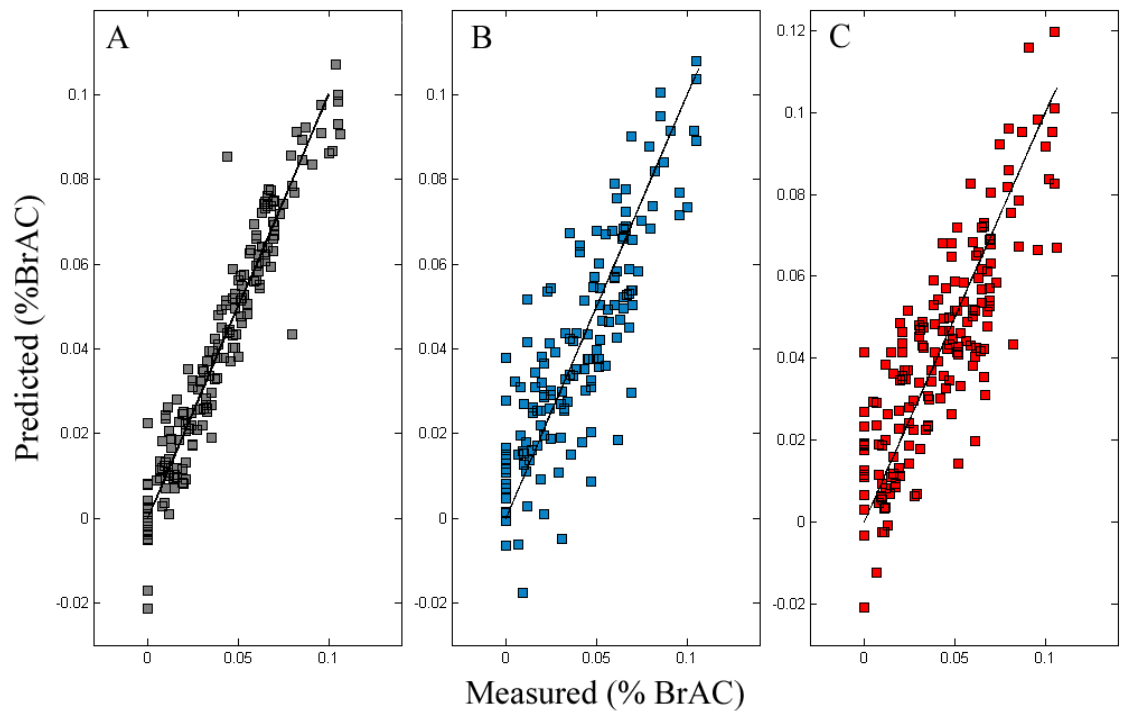


Figure 3.6: Global calibration plots of predicted % BrAC versus the actual % BrAC using leave-one-out cross validation based on **A:** PCA of MFC NIR image data: $r^2 = 0.987$ and $SEP = 0.0081$ % BrAC. **B:** Interferometer recorded speech for word “bit”: $r^2 = 0.852$ and $SEP = 0.0153$ % BrAC. **C:** Microphone recoded speech for word “boot”: $r^2 = 0.8432$ and $SEP = 0.0151$ % BrAC.

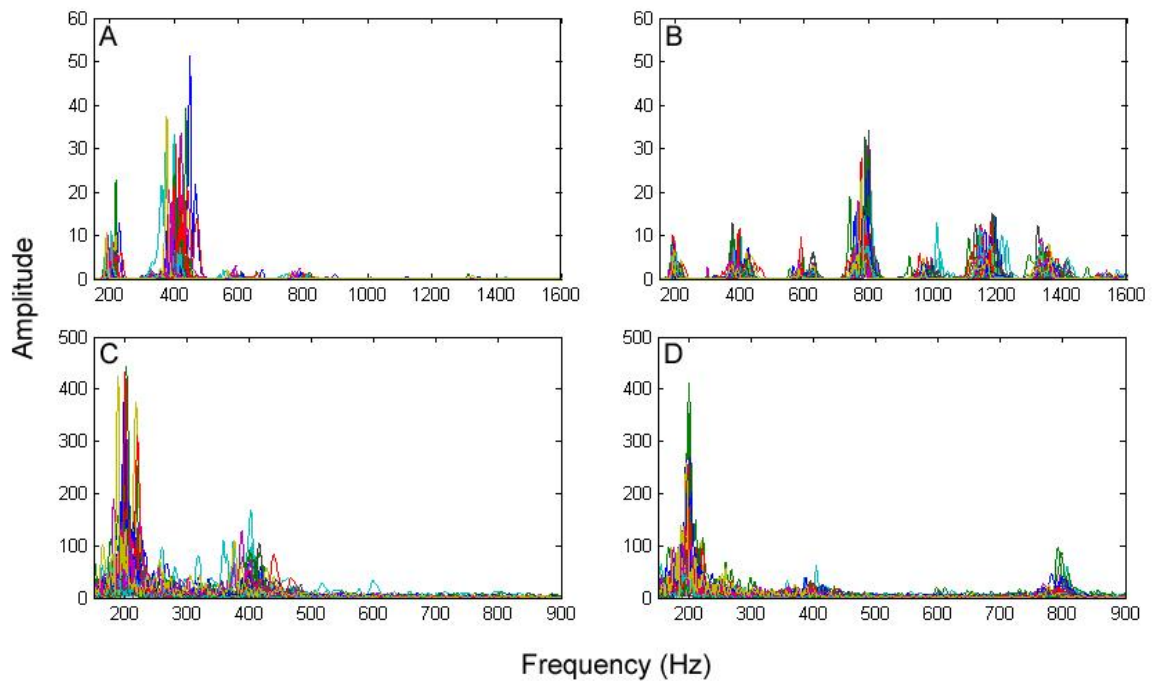


Figure 3.7: Speech spectra from patient 5 showing frequency differences between words and response of the interferometer compared to the microphone. A and B show the microphone-detected frequency spectra of words “boot” and “light”, respectively. C and D show the interferometer-detected frequency spectra of words “boot” and “light”, respectively.

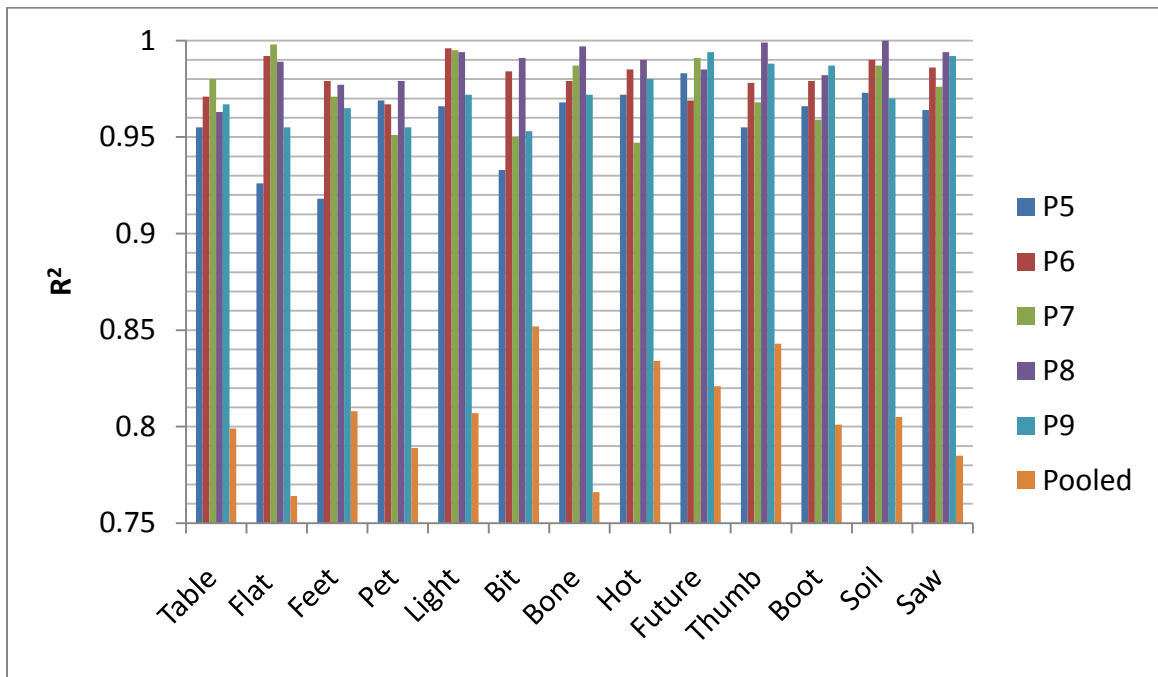


Figure 3.8: Comparison of leave-one-out cross-validation correlations for predicted versus actual %BrAC from analysis of interferometer-recorded speech. Individual correlations were found with $R^2 > 91\%$.

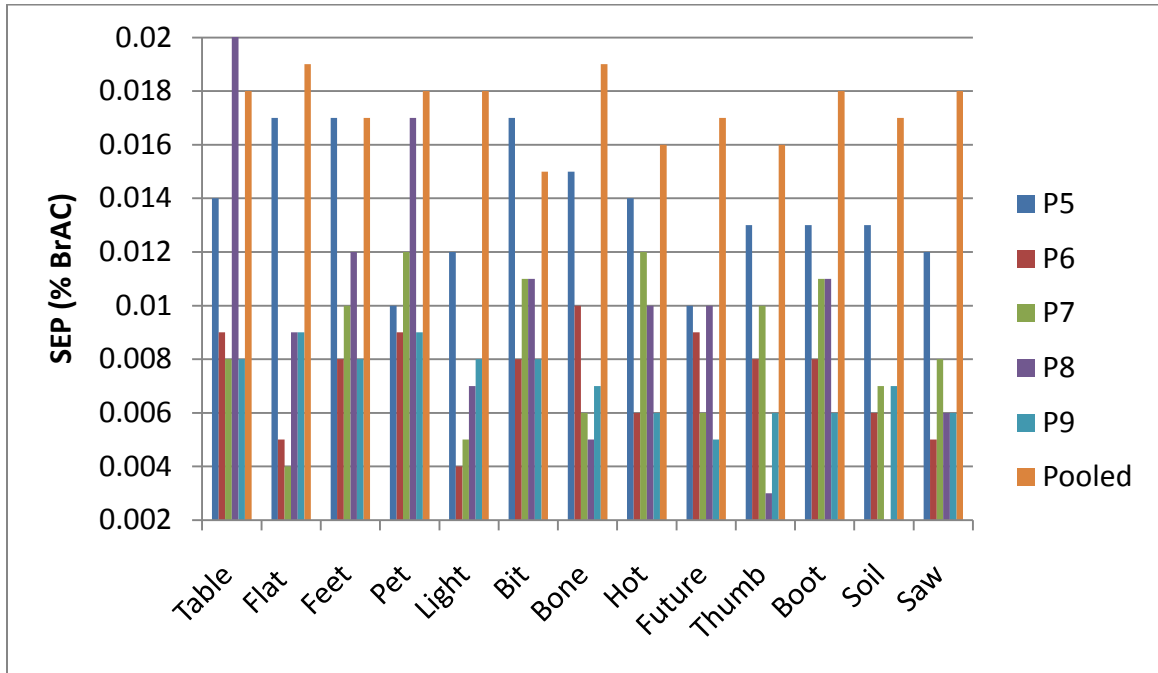


Figure 3.9: Comparison of SEP from leave-one-out cross-validation for predicted versus actual %BrAC analysis of interferometer-recorded speech. The SEP for individual measurements was < 0.020 % BrAC.

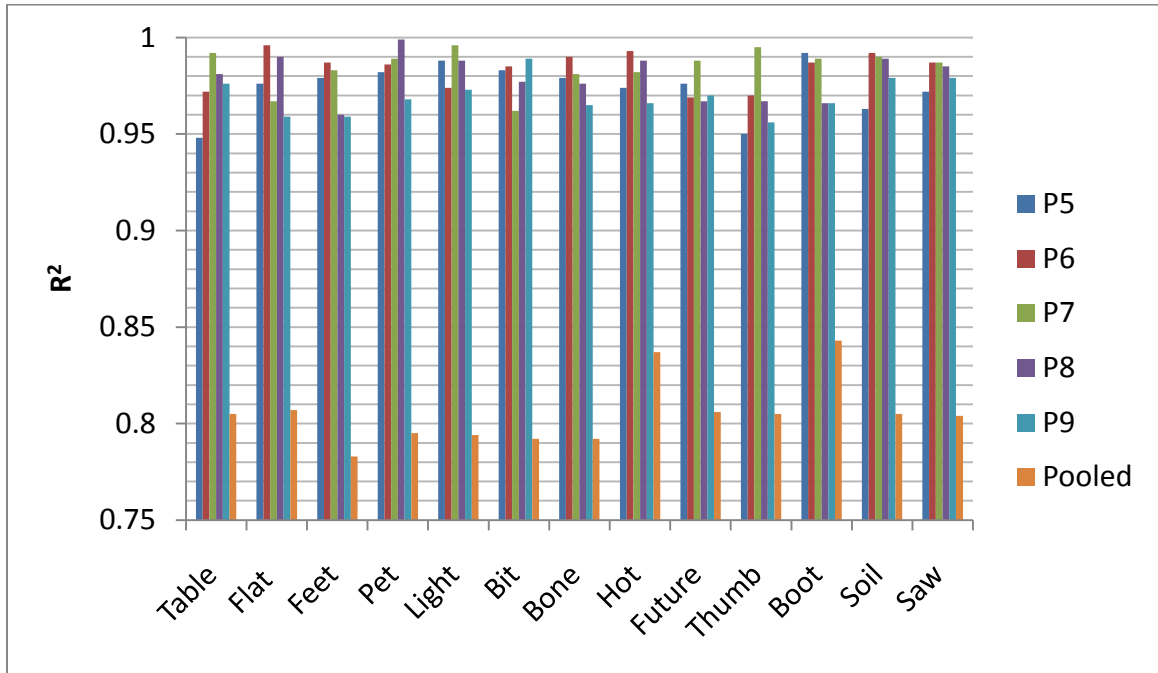


Figure 3.10: Comparison of leave-one-out cross-validation correlations for predicted versus actual %BrAC from analysis of microphone-recorded speech. Individual correlations were found with $R^2 > 94\%$.

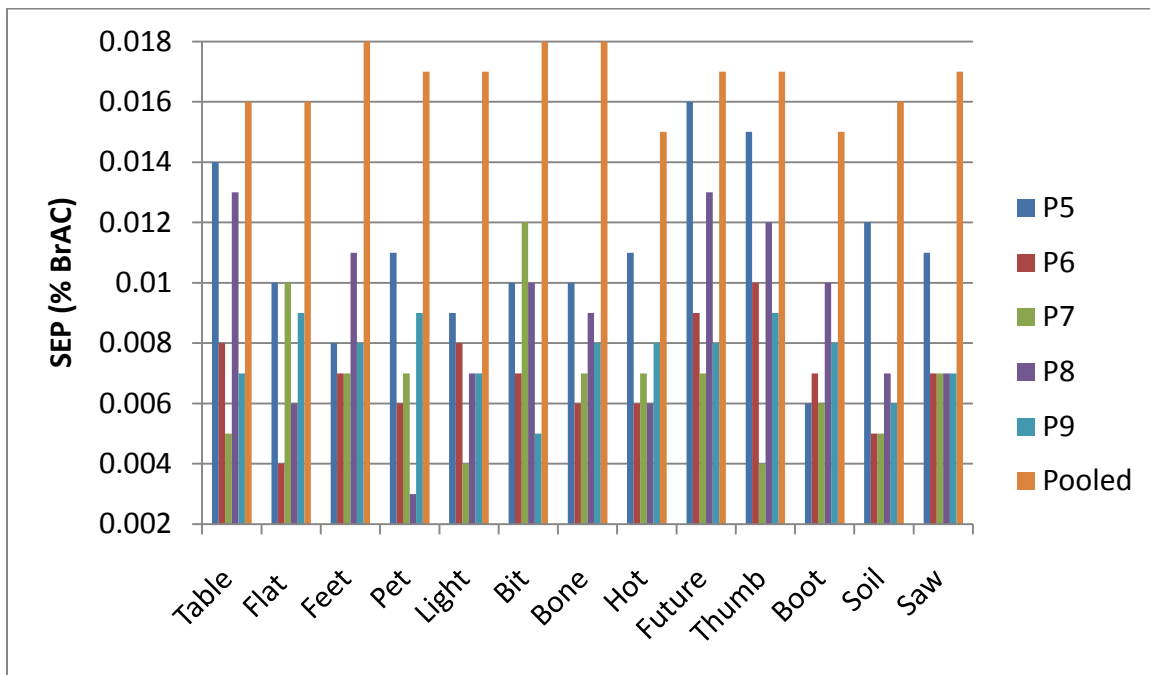


Figure 3.11: Comparison of SEP from leave-one-out cross-validation for predicted versus actual %BrAC analysis of microphone-recorded speech. The SEP for individual measurements was < 0.016 % BrAC.

Chapter Four-Bootstrap Error-adjusted Single-sample Technique Molecular Factor Computing (BEST-MFC) Near Infrared Hyperspectral Imaging: Classification of False Samples

Introduction

One problem with conventional alcohol monitoring modalities is that the subject is aware that he or she is being monitored. A noninvasive and surreptitious measurement of breath alcohol content (BrAC, known to be highly correlated to blood alcohol content ⁷³) is of interest because of the many problems regarding the existing real-time and more intrusive methods. The previous chapter detailed research on the use of molecular factor computing (MFC) Near-Infrared hyperspectral imaging for noninvasive and surreptitious monitoring of BrAC. Near-IR hyperspectral imaging is a useful tool for quantitative analysis of biological samples because it can provide rapid measurements with little or no sample preparation¹⁰⁰⁻¹⁰². MFC near-IR hyperspectral imaging is of interest for remote alcohol sensing because of the advantages associated with integrating sensing and processing (ISP). In ISP the chemometrics associated with post-processing are integrated into the sensing process. With MFC near-IR imaging, measurements of BrAC can be made without the subject being aware and yield an immediate result. The basic theory of MFC has been illustrated by Myrick et al. in UV-visible, NIR, and Raman spectroscopy¹⁰³⁻¹¹¹. The method of MFC was detailed in the previous chapter. In short, MFC filters are selected to model the positive and negative portions of a PC loading vector. The PC loading vectors used to predict an unknown (in this case, ethanol concentration) are calculated from a calibration set comprising a wide range of

concentrations. Therefore, the training set is used to define a pattern that the MFC filter system will detect. Thus the MFC filter system is a pattern recognition algorithm that is performed optically. Our laboratory has demonstrated the use of MFC filters for spectral encoding in several publications^{5,66,112}. The applicability of the MFC method is limited by the fact that it is a pattern recognition algorithm. Problems arise in pattern recognition when the pattern-recognition algorithm encounters a sample unlike any in the original calibration set. This is termed the false sample problem. The false sample problem is not an issue if the calibration set used to select the MFC filter contains every substance in the known universe. Figure 4.1 depicts the absorption regions from overtones and combinations of common functional groups in the near-IR region of the electromagnetic spectrum. Consider the construction of a MFC filter for ethanol detection. If a sample contains nothing but ethanol, theoretically, a perfect MFC filter for hyperspectral imaging could be composed from molecules containing all functional groups except –OH. All light would be blocked by the MFC filter except light that reveals ethanol concentration. The MFC filter would be immune to false samples as long as functional groups absorbed light completely and only where indicated in a spectra-structure correlation chart. However, in real life everything absorbs and scatters light to some degree at all wavelengths in the near-IR. Different chemicals just absorb light in slightly different amounts at all wavelengths. It is this property of differential absorption and scattering that creates a need for multivariate analysis and can lead to false-sample problems.

Hyperspectral images contain both spatial and spectral information. In MFC the output voltage of the detector is a single dimension that corresponds to analyte concentration. However, in MFC imaging each pixel output of the detector array

becomes a matrix of voltages corresponding to analyte concentration at many points in space. An erroneous prediction can be made if portions of the pixels in an MFC hyperspectral image comprise signals from false samples. To this end, it becomes necessary to assign statistical significance to the voltage output class. While there are many classification techniques available most are computationally demanding. One classification method is the Bootstrap error-adjusted single-sample technique (BEST). The BEST method is more computationally efficient than other techniques (e.g. Mahalanobis = $O(n^3)$, BEST = $O(n)$). However, because bootstrap replicates are calculated to estimate a population density, it becomes more difficult to perform on-line. One of the advantages of integrating sensing and processing is the speed of the method. In order to utilize the advantage of speed in ISP a fast validation method is needed.

The BEST-MFC method utilizes BEST distances calculated from a bootstrap distribution off-line. BEST multi-dimensional standard deviations (MSDs) are calculated for each point of a bootstrap distribution. The bootstrap MSDs are normalized to one BEST MSD and stored in a hash table. BEST distances are estimated for a new sample by calculating the dot product between the sample and hash table. The BEST hashing method is faster on a digital computer than the full BEST algorithm, illustrating utility for classification of MFC image data. The following research illustrates the application of the BEST-MFC hash method for classification of false sample data in MFC hyperspectral images.

Materials and Methods

The experimental and clinical trial procedures for the surreptitious remote sensing of blood alcohol concentration were described in the previous chapter.

MFC near-IR image data for a single subject were used in this study. The images were selected because the subject had facial hair. Image pixels comprising reflectance data corresponding to facial hair were used as false samples. All six filters used in our previous ethanol experiment were employed in this study and were as follows: polycarbonate (F1; Lexan, Plaskolite, Inc.), Acrylic (F2; Unbranded), polyvinyl chloride (F3; Unbranded, McMaster-Carr), combined gel filters CC20B and CC40G (F4; Kodak), gel filter Pale Yellow (F5; Kodak A2), and polymethyl-methacrylate (F6; Optix, Plaskolite, Inc.).

Image data corresponding to subject P9 were loaded into MatLab (R2008a) where they created a matrix of intensity values with a size of 720x480x3 (rows by columns by layers). The layers corresponded to red, green, and blue colors spaced such that the combination of the intensities at a pixel produced 1 of 256 possible colors. For each measurement corresponding to an ethanol concentration, six images (one for each MFC filter) were acquired. The total data for each concentration measurement comprised a matrix with a size of 720x480x18 (rows by columns by dimensions). The dimensions correspond to the red, blue, and green color data for each MFC filter image. In total, 37 concentration measurements of subject P9 were analyzed. Each image pixel represented a spectrum containing 18 predictor variables (dimensions). In total, there were 12,787,200 spectra (720 x 480 pixels x 37 measurements) that could have been examined. Because of the high volume of data to be analyzed a small subset was used for this study. Each image was sampled at different pixels to generate 2000 skin (intensity values to correspond to BAC) and 2000 facial hair (false sample) spectra. The spectral sets were averaged to yield 500 spectra for each class.

The skin spectra were used to construct a BEST hash table comprising 2000 points. For each concentration measurement BEST multidimensional standard deviations (MSDs) were estimated for both the skin and facial hair spectra. MSDs greater than 3 are considered to be from false samples.

Results and Discussion

Figure 4.2 illustrates the spectra for pixels corresponding to skin and facial hair. The figure demonstrates the differences in reflectance between areas of skin and hair. The signals from facial hair were at higher intensities than those from skin, illustrating the high reflectivity of hair. Inclusion of pixel data illustrated in Figure 4.2 corresponding to facial hair into the prediction model will result in one of two possible responses. Either the prediction model will predict a reasonable estimate of BrAC or an estimate that is obviously wrong. When image data from facial hair was included into the prediction model there were three instances where a reasonable prediction was given. Table 4.1 lists the actual percent BrAC values versus the predicted values. In these three cases the subject would have been predicted well above or below the actual BrAC level.

The BEST-MFC method operates on the assumption that like samples cluster together in multidimensional space. Training set samples corresponding to skin were used to calculate a standardized MSD contour based off of BEST distances. The points comprising the contour were stored in a hash table. BEST MSDs were estimated for each spectrum from both skin and facial hair regions by comparing spectra from each to the hash table. Figure 4.3 illustrates a 3 MSD contour in three dimensions calculated using the BEST-MFC hash method. It is evident from the figure that the skin and facial hair spectra are well separated. Table 4.2 indicates the classification statistics for both skin

and facial hair using the full BEST algorithm and the BEST hash metric, respectively. The statistics listed in Tables 4.2 were calculated using all 18 dimensions (6 filters x 3 color layers). In all cases facial hair was always identified as a false sample.

To determine if the performance of the classification model changes with respect to the number of filters used, the BEST-hash method was performed on the image data with varying dimensions. Table 4.3 lists the specificity and sensitivity for identifying 500 false sample (facial hair) spectra from 500 skin spectra using a 2000 variable BEST table for a given number of dimensions. While skin and facial hair spectra were well separated in multidimensional space, Table 5.3 indicates that the performance of the BEST-MFC metric degrades somewhat as the number of dimensions increases with respect to specificity. The change in specificity indicates that some skin spectra are identified as false samples. Because the BEST MSD between skin and facial hair spectra is so large the standard deviation cutoff could be increased to account for more variation in the skin spectra. In the previous chapter it was found that for most cases all filter data were needed to predict BrAC. For qualification, however, only a few filters are needed to accurately identify facial hair as a false sample due to the variation between spectra. It is conceivable that image data from a small number of filters could be used to determine whether to load all image data into a calibration. The same filters used to identify false samples can also be used to determine which image pixels to use in a prediction. Figure 4.4 illustrates the estimated BEST MSDs calculated with image data from 3 filters (F2, F4, and F5). The figure indicates only the subject's lower left face because of IRB regulations on patient privacy. The subjects' facial hair is clearly identifiable in the

figure. Use of the estimated BEST MSDs to classify pixel regions in the image permits automation of the system.

With BEST-MFC the statistical significance of the classification is also output with the measurement value. The BEST-MFC method is a much faster method of classification than the full BEST algorithm on a digital computer. The results in Table 4.2 required on average 1.8 seconds to complete with the BEST hash method compared to an average of 18 minutes for the full BEST algorithm. Table 4.3 also indicates the difference in run times for the BEST hash method for a given number of dimensions.

The BEST hash method also permits the identification of inactive pixels. Software for correction of dead pixels is commonly included with most modern near-IR imaging equipment. Software corrections generally use an interpolation process to replace dead pixels. The replacement procedure is automatic and is carried out without the user being informed. This can become problematic when many different samples are imaged. When there are a limited number of pixels on the detector array available for each analyte, interpolated values can cause erroneous estimates of some analytes. Figure 4.5 illustrates a group of bad pixels on the InSb focal plane array used in the study. The standard deviation image clearly indicates the bad pixel region. Removal of the data from bad pixel regions is necessary for accurate analysis of multiple samples.

Conclusion

BEST-MFC near-IR hyperspectral imaging has been investigated as a method of identifying false samples. Near-IR hyperspectral imaging offers the ability for noninvasive and surreptitious monitoring of samples. MFC near-IR hyperspectral

imaging has the advantage of speed by integrating the sensing and processing in the physical design of the collection process. BEST-MFC offers the advantage of applying statistical significance to the class output from the detector at speeds faster than the full BEST algorithm. The BEST-MFC method classified 1000 spectra in only 1.8 s which required over 18 minutes with the full BEST method. In all cases the accuracy, precision, sensitivity, and specificity were equal to or better than the full BEST method. The fast false sample detection ability gives BEST-MFC hyperspectral imaging utility as an automated sensor system.

Chapter Four Tables

Table 4.1: Erroneous predictions resulting from inclusion of facial hair spectra into the prediction model.

True concentration % BrAC	False Prediction % BrAC
0.060	0.010
0.047	0.088
0.019	0.094

Table 4.2: Performance statistics for BEST classification of skin and facial hair spectra using the full BEST algorithm and the BEST hash method. The BEST-MFC method classified 500 skin and 500 facial hair spectra in only 1.8 s which required 18 minutes for the full BEST algorithm.

Method	Full BEST		BEST-MFC	
	Skin	Hair	Skin	Hair
Accuracy (%)	99.3	100	99.8	100
Precision (%)	100	100	100	100
Correct	493	500	498	500
Specificity	1	1	1	1
Sensitivity	0.986	1	0.996	1

Table 4.3: Performance results of the BEST-MFC method for identifying 500 false sample spectra from 500 skin spectra with varying degrees of dimensionality.

# Dimensions	3	5	7	9	11	15	18
Specificity	1	1	0.996	0.996	0.996	0.996	0.996
Sensitivity	1	1	1	1	1	1	1
Run Time (s)	0.401	0.500	0.617	0.791	0.939	1.291	1.502

Chapter Four Figures

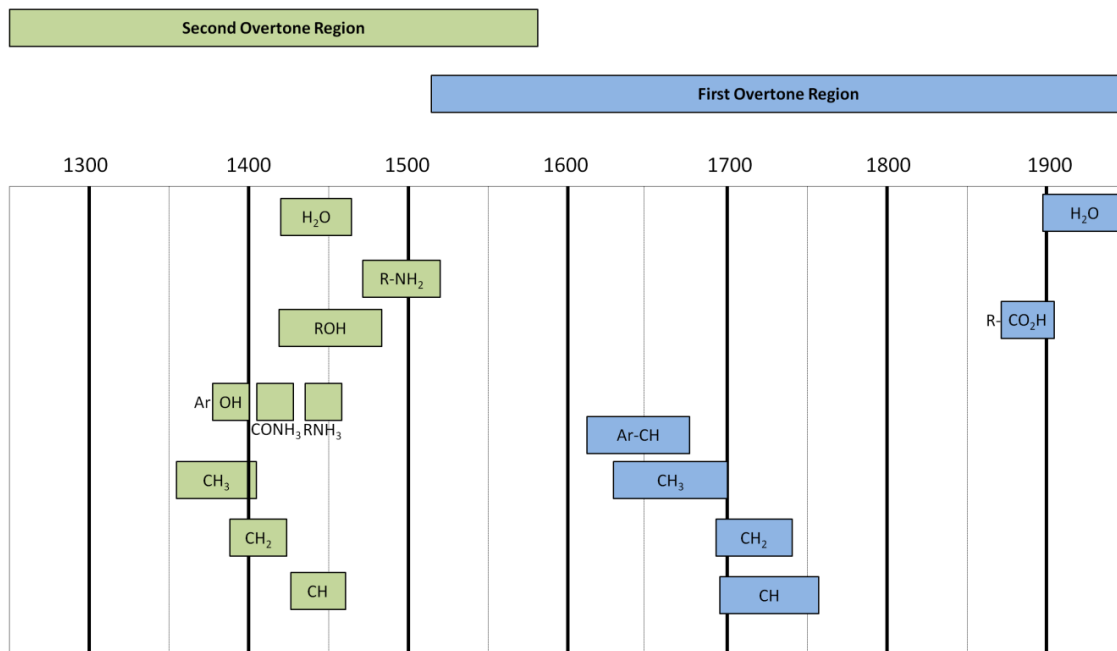


Figure 4.1: Near-infrared correlation chart for common functional groups. This chart is only an approximate guide for near-IR band assignment by functional group.

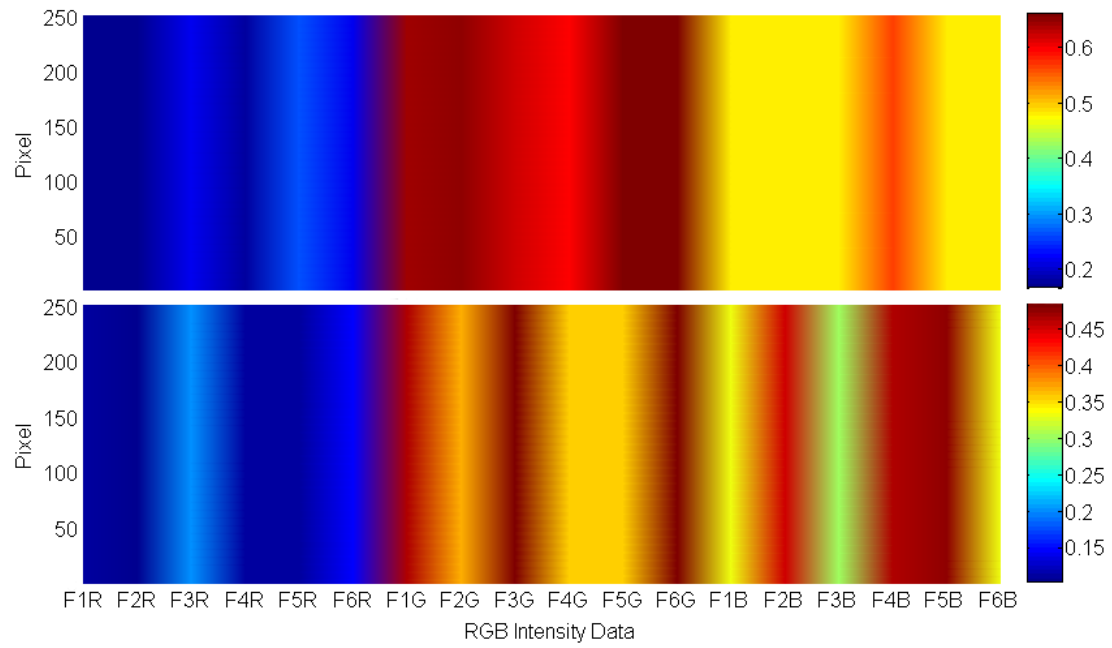


Figure 4.2: Reflectance data for selected pixel regions comprising skin (bottom) and facial hair (top). Reflectance intensity is indicated by color, increasing from blue (low intensity) to red (high intensity).

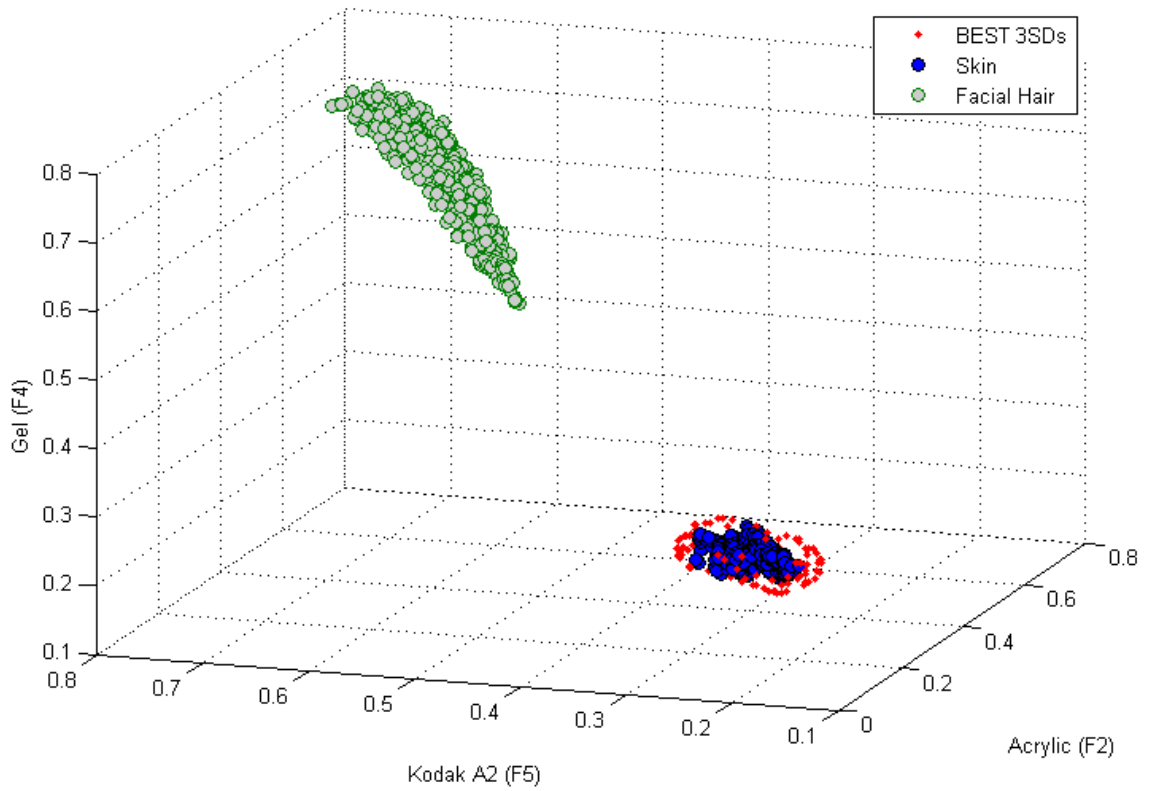


Figure 4.3: Estimated BEST MSD contour at 3 SDs for filters F2, F4, and F5. The BEST contour represents 100 points taken from a 2000 point BEST Hashtable.

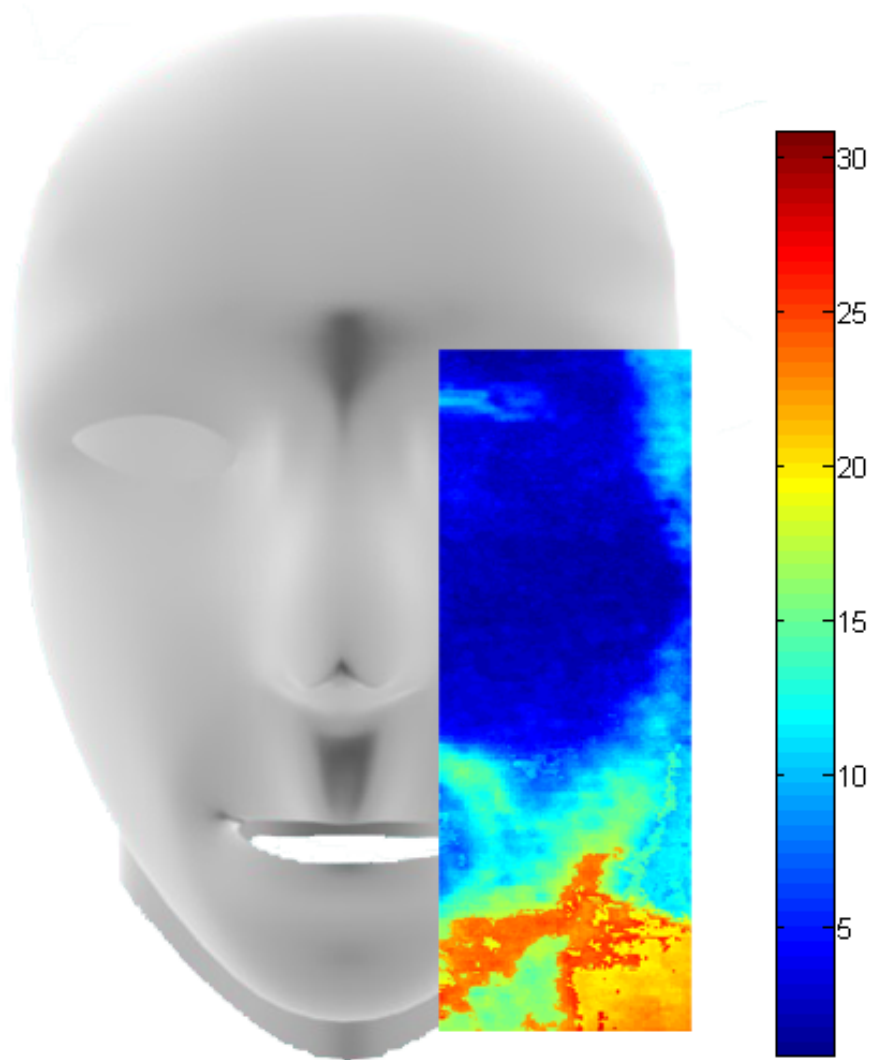


Figure 4.4: Estimated BEST MSD plot of subject 9. Subjects' full face is not shown due to IRB- approved protocols regarding patient privacy (approval number 07-0417-F1V). BEST MSDs are indicated by color increasing from blue (low MSD) to red (high MSD). MSDs greater than 3 are considered to be different from the training set.

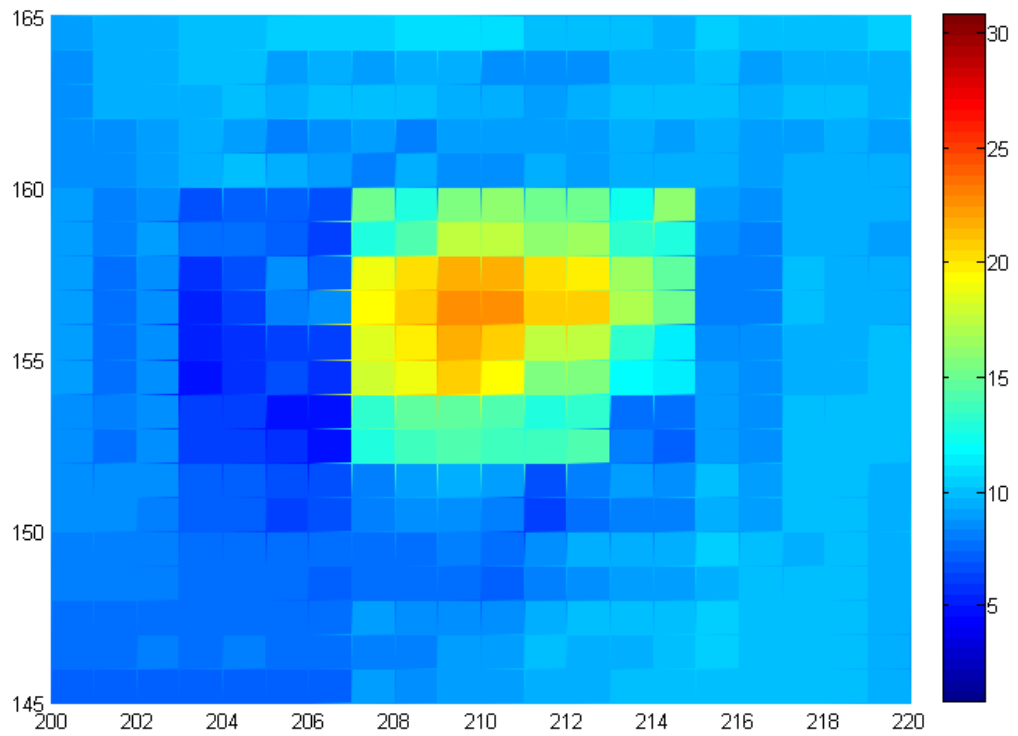


Figure 4.5: Estimated BEST standard deviation plot illustrating a bad pixel area on the InSb focal plane array. BEST MSDs are indicated by color increasing from blue (low MSD) to red (high MSD).

Section IV: Astrobiology

Astrobiology is the study of the origin, distribution and evolution of life in the Universe¹¹³. As a multidisciplinary science astrobiology, combines fields such as chemistry, biology, computer science, geology and many others. A key aspect of the NASA astrobiology program is the detection of extinct or extant life on Mars¹¹⁴. NASA has sent several robotic probes¹¹⁵⁻¹¹⁷ and is in the process of planning future missions¹¹⁸ to Mars with the purpose of detecting past or present life or searching for suitable environments for life. Mars is typically at the center of astrobiology research because it is thought to have had a similar geological history to Earth¹¹⁹⁻¹²⁰. Organisms, termed extremophiles can be found on Earth in remote and harsh environments and are some of the oldest life forms known. If life exists or ever did exist on Mars it is likely that it would resemble the extremophilic organisms found on Earth. The geological and environmental strategies that extremophiles must take to adapt to such harsh planetary environments make them useful models to derive remote detection schemes.

Optical spectroscopy is well suited for the identification of both geo- and biomarkers associated with extremophilic organisms. Ultra violet (UV), visible and near-Infrared (IR) methods coupled with chemometric techniques can be used to identify complex organic and inorganic compounds associated with life. Near-IR spectroscopy is a useful tool for the identification of biological samples because it can penetrate samples well and excellent signal-to-noise ratios can be obtained¹²¹⁻¹²³. Raman spectroscopy has been previously used for the identification of astrobiologically significant organic and inorganic compounds¹²⁴⁻¹²⁶. These methods are suitable for sensing in remote environments because sample preparation and manipulation is not needed.

Raman spectroscopy is considered to be a viable tool for the search of extraterrestrial life. However, Raman scatter is very weak ($\sim 10^{-9}$ photons may be Raman scattered) making measurements of low concentrations more difficult compared to near-IR (one in 10 photons may be scattered). Many of the drawbacks of optical techniques are in the instrument design. Optics and moving parts can be problematic in remote sensing applications. Any movement of the internal spectrometer parts can have a drastic effect on the resulting spectra. An ideal spectrometer for the identification of bio- and geomarkers associated with extraterrestrial life would be robust, small and lightweight.

The following chapters describe research conducted with a novel solid state spectral imager (SSSI). The SSSI is a small and lightweight imager that is expected to use laser diodes at wavelengths ranging from UV to near-IR. The use of a pseudorandom pulse sequence allows for multiple diode illumination which results in an increased signal-to-noise ratio. The SSSI has no moving parts or optics normally associated with traditional spectrometers and imagers, making it a useful choice for remote sensing applications. Chapter six describes research for the development of an ISP routine for the SSSI. In this chapter copper sulfate solution was quantified based on the voltage output of the detector. The ISP algorithms described in the following chapters were based on a training set of discrete samples. In chapter seven an on-line BEST-ISP method was investigated for the determination of samples not included in the original training set (false samples). The final chapter is a hypothetical study and the author fully admits that the line between reality and fiction has been closely drawn. However, the reader should keep in mind that the theoretical approach investigated for the exploration of other worlds

is based on real and developing technologies and remember that science is only limited by the imagination.

Chapter 5- Complementary Integrated Sensing and Processing Solid State Spectroscopy (CRISP-SSSI) for Rapid Detection of Copper Sulfate

Introduction

Copper sulfate pentahydrate ($\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$) is commonly used in agriculture but has a broad range of other non-agricultural uses. In the United States $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ is used for bacteria and fungus control in food crops. Copper is toxic to microorganisms because of its ability to denature cellular proteins and deactivate enzyme systems¹²⁷. $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ is found naturally in the soil however, accumulation of toxic levels is problematic because it is difficult to remove. Furthermore, the accumulation of copper in soils may be harmful to microbes, which have an active role in nutrient cycling and decomposition of organic matter¹²⁸. The toxicity of copper to cyanobacteria¹²⁹ and other organic life on Earth is likely to exist on other planets as well.

During the Phoenix mission on Mars, perchlorates were discovered in the Martian soil. This recent discover led to a debate of whether it was possible for organic life to exist under such conditions¹³⁰⁻¹³¹. Perchlorate is a strong oxidizer and is used as a main ingredient in some rocket fuels. In high concentrations, perchlorates are harmful to organic life due to their oxidizing nature. However, perchlorates are also found naturally in low concentrations on Earth in places such as Chile's Atacama Desert, where extremophilic organisms are also known to exist.

Both copper and perchlorate are toxic to many microorganisms however; there are some that have the ability to tolerate low levels of these substances¹³²⁻¹³³. For both copper and perchlorate, the concentration in the surrounding environment will likely

determine whether organisms can be found. Many experiments that are designed to search for extant and extinct life intrinsically look for biosignatures (molecules associated with UV pigments, etc.). If the environment is not likely to support life then it may be useful to save valuable resources time by searching for life in places where the chances are considerably better. A method of rapid identification and quantification of toxic substances would allow a rover to spend more time searching for life in places where it is more apt to find it.

Solid State Spectroscopy The Solid-State Spectral Imager (SSSI) is a new device for acquiring hyperspectral images and spectroscopic data. A light emitting diode array is used to acquire ultraviolet (UV), visible and near-IR data of biological and geological samples. The SSSI comprises three main parts: a light emitting diode array, an on-board processor and a single photodetector. The use of a single photodetector is possible because of the unique data-analysis algorithm. The SSSI utilizes an orthogonal pseudo-random (e.g., a Hadamard) pulsing sequence¹³⁴. The Hadamard pulse sequence progresses stepwise in such a way that the signal from a single detector can be demultiplexed to determine the contribution of each LED wavelength to the total observed reflection spectrum. Hadamard encoding has the benefit of multiple (parallel) diode illumination rather than illuminating with diodes sequentially. Multiple diode illumination allows more energy to reach the detector increasing the signal. The Hadamard weighting provides equal covariance and duty cycles on all diodes, reducing premature hardware failures. Figure 5.1 illustrates the utility of Hadamard encoding for a 3-diode system. In Figure 5.1, each diode can be turned on and off separately, producing three equations with three unknowns representing the contribution of each diode

wavelength to a sample spectrum. The equations could be solved with brute force, but as the number of diodes increases the computational burden of solving a system of equations increases sharply. The Hadamard transform (HT) is an elegant solution to the problem of increasing complexity.

The SSSI is unique because it does not require moving parts or lenses so it can be reduced in size very easily (see Figure 5.2). The SSSI is rugged, giving it the ability to operate in extreme environments¹³⁵ making it a useful choice for the study of extremophilic organisms in remote places like Mars¹³⁶. In remote locations, the sensor system needs to have the ability to make decisions on its own. One method of accomplishing this is by integrating the sensing and processing directly into the detector. Integrated sensing and processing (ISP) incorporates information about the sample directly into the physical design of the data collection process. In traditional spectroscopy, sampling is performed to obtain data which is then analyzed to reveal information about the sample. The statistics gained from the analysis can be applied to future data sets but computation will always follow sampling. The computational processes following sampling in traditional methods are applied with ISP during the sampling process at the speed of light. Incorporating chemometrics into the detection process virtually eliminates the computations that normally follow the data acquisition process. This technique works well for situations where there is more than enough time for calibration before data collection, but limited time after. Advantages of using ISP include a decrease in computational and data storage demands by integrating the sensing and processing directly into the transducer (the integrated sensing and processing, or ISP, advantage), the Fellgett advantage (acquisition of all wavelengths simultaneously permits

higher signal-to-noise ratio for a given acquisition time) and the Jacquinet advantage (the absence of monochromator slits allows higher optical throughput). The SSSI is capable of operating in a form of ISP, complementary randomized integrated sensing and processing (CRISP) scanning mode. In this mode, information that is gained from Hadamard scanning is applied to produce a real-time analysis technique.

Theory

In CRISP scanning mode, factor-based analysis techniques such as principal component analysis (PCA) are used to maximize the sample variation in a defined bandpass. PCA is a common tool used for preprocessing multivariate data, but PCA can be computationally expensive, requiring an $O(n^3)$ matrix factorization¹⁵. In PCA a linear transformation of the data from one coordinate system to another of reduced dimensionality is performed. Calculation of PCs can be accomplished by performing the singular value decomposition (SVD) on a matrix, \mathbf{A} (the spectra) of dimension $m \times n$, according to Equation 5.1:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (5.1)$$

where \mathbf{U} ($m \times m$) is an orthonormal matrix representing PC scores, \mathbf{S} ($m \times n$) is a matrix of singular values, \mathbf{V}^T ($n \times n$) is an orthogonal matrix of eigenvalues (loadings). The dimensions of the matrices are dependent on the number of samples, m and the number of predictor variables, n . The decomposition of \mathbf{A} is calculated such that the maximum variation of the data set is found in the first PC. Each successive PC contains a lesser percent of the variation and is orthogonal to the previous one. In a pure two-component system only a single PC is needed to describe the sample. However, spectroscopic data is

not as straightforward. Interferences, detector noise, spectral overlap and other inconsistencies complicate the sample and the number of PCs needed to describe the variation increases. Generally, the number of PCs needed to describe a sample is taken to be less than half of m and in practice is usually far less.

In PCA, spectra are transformed into spectral scores. Each signal, a recorded with a narrow bandpass is weighted at each wavelength, λ with a coefficient, f over n wavelengths.

$$score = f_1 a_{\lambda_1} + f_2 a_{\lambda_2} + f_3 a_{\lambda_3} + \dots + f_n a_{\lambda_n} \quad (5.2)$$

In PCA, the factors calculated by SVD are based on the linear variation in the spectral data set. In CRISP optical representations of the factors (f in Eq. 5.2) are created to weight the signal (a in Eq. 5.2) by programming pulse sequences to match the loading vectors generated in PCA. However, LEDs can only be in one of two states, represented as a binary one (on) or zero (off). In equation 5.2, f_i is the loading coefficient between zero and one quantifying the contribution of the i^{th} wavelength (LED) to the *score*. For a CRISP sequence to match the loading vector, the on and off states with varying duty cycles must be integrated over time so that the summed values correspond to a factor score. In this implementation of CRISP, more pulses are used per integration frame at wavelengths with higher loading values (see Figure 5.3). The CRISP modulation bit rate (bandwidth) controls the intensity resolution of the factor result.

CRISP pulse sequences are randomized to produce as little covariance between the sequences as possible. The length of sequences used for each loading depends on the desired resolution. At the lowest resolution, a CRISP sequence would encode all loading

values as a binary one or zero. As the number of pulses and integration frames is increased, the intensity resolution can become quite high, allowing the loading coefficients from PCA to be closely simulated. Together, the encoding of spatial and intensity information on the LEDs permits the simultaneous identification, quantification, and location of specific analytes.

Materials and Methods

Materials. Copper sulfate pentahydrate ($\text{CuSO}_4 \cdot 5 \text{H}_2\text{O}$) (J.T. Baker Chemical co.) was massed and diluted to 500 mL to create a series of solutions ranging from 480 to 7604 ppm. The solutions were scanned with the SSSI in Hadamard scanning mode. A cuvette was constructed in-house with a path length of 10cm.

Data Collection. Solutions were scanned in triplicate and in random order. The SSSI was set to acquire 50 replicate scans per each scan acquired, which resulted in 150 scans for each solution. All solutions were scanned in ambient light and in random order. The prototype SSSI was controlled via a serial connection to a laptop computer. A graphical user interface (GUI) was created in MatLab 7.0.1 (The Mathworks Company, Natick, MA, USA) to send commands to the SSSI CPU and transfer all data. Data were imported into MatLab for analysis. The spectra were scatter corrected and replicates were averaged before PCA was performed. Leave-one-out cross validation was used to validate the multivariate PC regression model. The loading vectors corresponding to the PCs used in the regression model were used to build CRISP sequences.

CRISP Sequence Construction. Loadings corresponding to the three PCs used in the validation set were required to derive the CRISP codes. The positive and negative loading data was separated. The absolute value of the negative loading data values were used because optical intensities cannot be negative. The loading values were normalized to one by dividing each loading vector by the maximum vector value. Because CRISP can only be implemented with simple 0- or 1-emission pulses the detector signal must be integrated for one pulse sequence length. However, the length of a CRISP code is determined by the number of bits of intensity resolution used. For a simple CRISP code, the intensity resolution is set to one. The normalized loading vectors were multiplied by one and rounded to the nearest whole number. This produced an intensity resolution of one. Table 5.1 illustrates how the CRISP sequence resolution is constructed. Clearly, an iterative approach could be used to generate a perfectly matched CRISP sequence to a loading vector, however the sequence may require more bits than are available in the processors memory. A binary representation matrix was constructed from each resolution vector, which resulted in a 1x25 matrix corresponding to LED on and off states for each positive and negative loading used. In total, three PCs were used to validate the training set which resulted in six 1x25 CRISP sequences. For higher resolution sequences the on and off states were randomized so that covariance within the sequences was minimized. Two separate CRISP codes were constructed at two different resolutions, 1 and 10 (R1 and R10, respectively). The CRISP codes for R10 were constructed from the loadings corresponding to the top 3 PCs. This resulted in six sequences that were 10 x 25 each. In total, CRISP for R10 was a 60x25 matrix.

Results and Discussion

Copper sulfate solutions were investigated with the SSSI in the visible region of the electromagnetic spectrum. Figure 5.4 is a depiction of the mean transmittance spectra from Hadamard scans. The mean transmission spectra in Figure 5.4 are characteristic of various concentrations of CuSO_4 solutions because the transmitted light is representative of the color of the solution. The spectra show a high transmittance for the solutions at 565 nm (green LED), which is due to the cuvette used in these experiments. The cuvette was constructed from glass that had a slight green tint. While the solutions should characteristically transmit blue light, the solutions in the cuvette transmitted blue-green light. The change in concentration adhered closely to Beer's law.

Principal component analysis (PCA) was performed on the transmittance spectra to determine the contribution of each diode in quantifying CuSO_4 . Three PCs were needed to cross validate the data. Figure 5.5A indicates the leave-one-out cross validation plot of measured CuSO_4 versus predicted CuSO_4 for the Hadamard scans, $r^2 = 0.989$, $\text{SEP} = 423$ ppm.

The loadings corresponding to the PCs used in cross validation were used in construction of the CRISP pulse sequences. The CRISP sequences were prepared with intensity resolutions of one and ten. Figure 5.6 and 5.7 indicates the pulse sequences for the positive and negative loading sequences for R1 and R10, respectively. For R1 the resolution was as low as possible. A low intensity resolution was investigated to save memory space on the processor. As the intensity resolution increases the amount of space available to store the codes can decrease drastically. As described previously loading values for the CRISP pulse sequences can only be set as binary one or zero (i.e.

on or off). Some values were set to zero when rounding the normalized loading vectors which. This was not considered problematic because the low loading values represent small contributions to the correlation of CuSO₄. Figure 5.5B and C are leave-one-out cross validation plots of measured CuSO₄ versus predicted CuSO₄ for CRISP scans R1, $r^2 = 0.997$, SEP = 296 ppm and R10, $r^2 = 0.992$, SEP = 416 ppm. Comparison between the Hadamard and CRISP correlation data support a clear ISP advantage. The CRISP pulse sequences at R1 and R10 both outperformed the conventional method; however, R1 unexpectedly outperformed R10. The outperformance of the higher resolution CRISP scans by the lower resolution may be due to the simplicity of the system.

Figures of Merit. The signal-to-noise ratio (S/N) was calculated as the mean intensity of the signal divided by the standard deviation of replicate measurements. The S/N ratio for the Hadamard scans was 20.5 but was much higher for CRISP scans at 60.4 and 45.0 for resolution equal to 1 and 10 respectively indicating a multiplex advantage. The limits of detection (LOD) were calculated by assuming a linear scale between the lowest concentration data point and the zero-concentration baseline and taking the concentration value corresponding to an S/N ratio of three. The LOD for the CRISP method was much lower than the Hadamard (Hadamard = 69 ppm, CRISP = 19.8 and 25.3 ppm for R1 and R10, respectively) also supporting a clear ISP advantage.

Conclusion

Complementary randomized integrated sensing and processing has been explored as a real-time analysis for the quantification of copper sulfate in solution. A SSSI with preloaded codes could be used to identify specific analytes in remote environments. The SSSI is a low cost and rugged spectrometer. The SSSI has potential uses in remote

environments such as Mars where it could be deployed in mass to identify areas of interest. The CRISP scanning method has been shown to outperform Hadamard scanning indicating ISP, Fellgett, and Jacquinot advantages.

Chapter Five Tables

Table 5.1: CRISP resolution calculation for the first seven terms of a positive loading vector. The desired CRISP resolution is acquired by first normalizing the loading coefficients. The normalized loading values are multiplied by the desired resolution then rounded. The length of the CRISP code is determined by the largest rounded variable.

Loading Values	0.4378	0.6369	0.1798	0.0196	0.9936	0.5974	0.2855
Normalized Loading Values	0.4406	0.6410	0.1810	0.0198	1.0000	0.6012	0.2873
Intensity Resolution (x1)	0.4406	0.6410	0.1810	0.0198	1.0000	0.6012	0.2873
Rounded Values	0	1	0	0	1	1	0

Chapter Five Figures

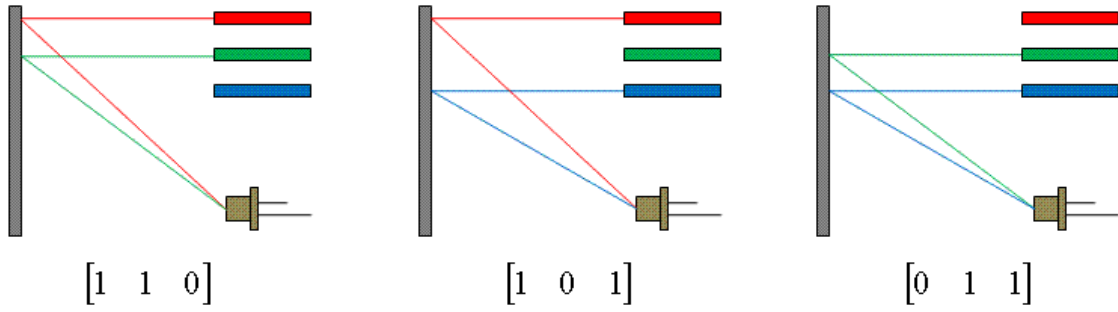


Figure 5.1: Hadamard encoding allows for multiple diode illumination and detection using a single diode. Multiple diode illumination has the benefit of increased signal at the detector.

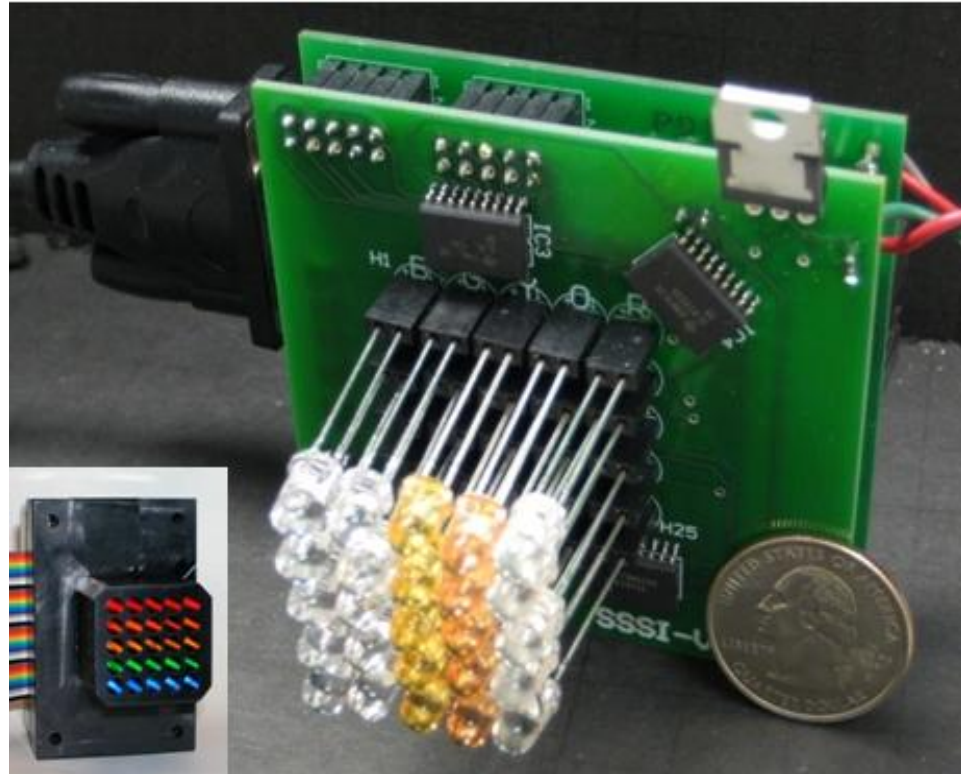


Figure 5.2: The current SSSI prototype comprises a 5x5 LED array in the visible wavelength region and measures 70x70x35mm. Inset is a SSSI with a collimator installed over the diodes. The collimator provides better spatial resolution.

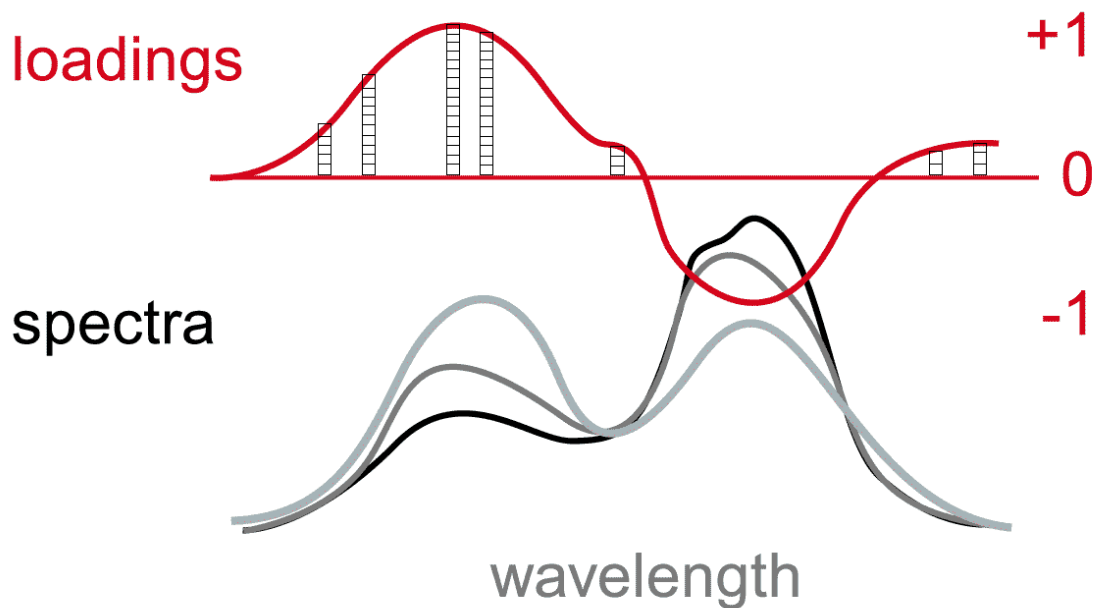


Figure 5.3: CRISP is employed by varying the duty cycle of the pulses (pulses represented by blocks in loadings graph) at each diode wavelength, so that in each integration frame, the most pulses are delivered at the wavelengths where the most light is needed to make the integration sum to the analyte value sought.

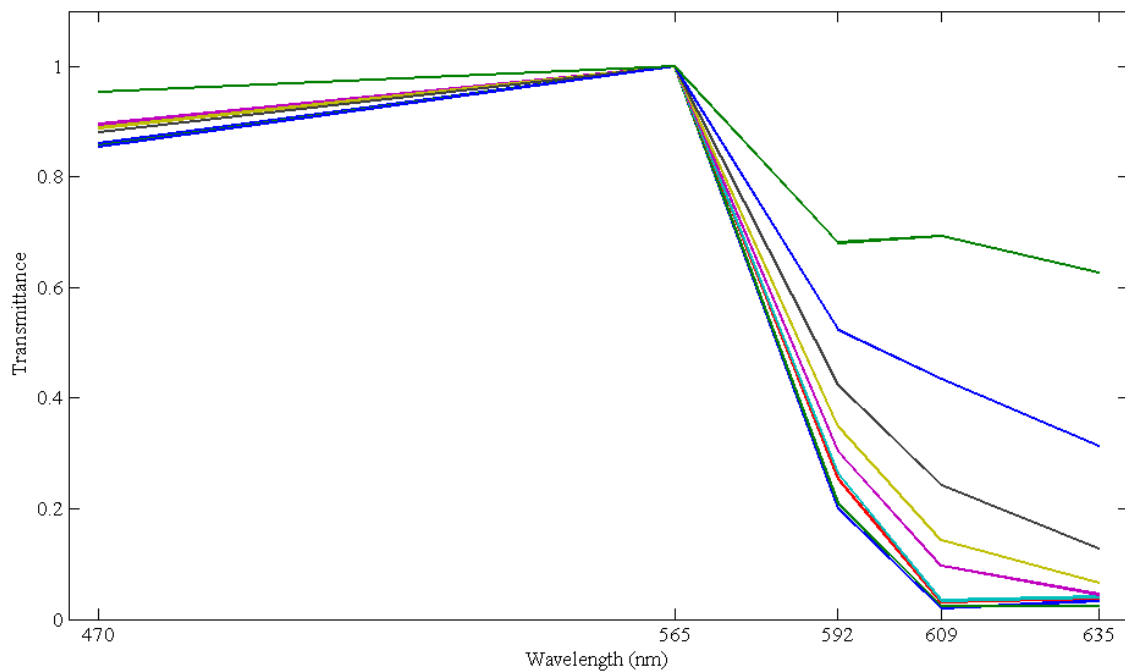


Figure 5.4: Mean transmission spectra of copper sulfate solutions. Each wavelength intensity value is the mean response of five LEDs.

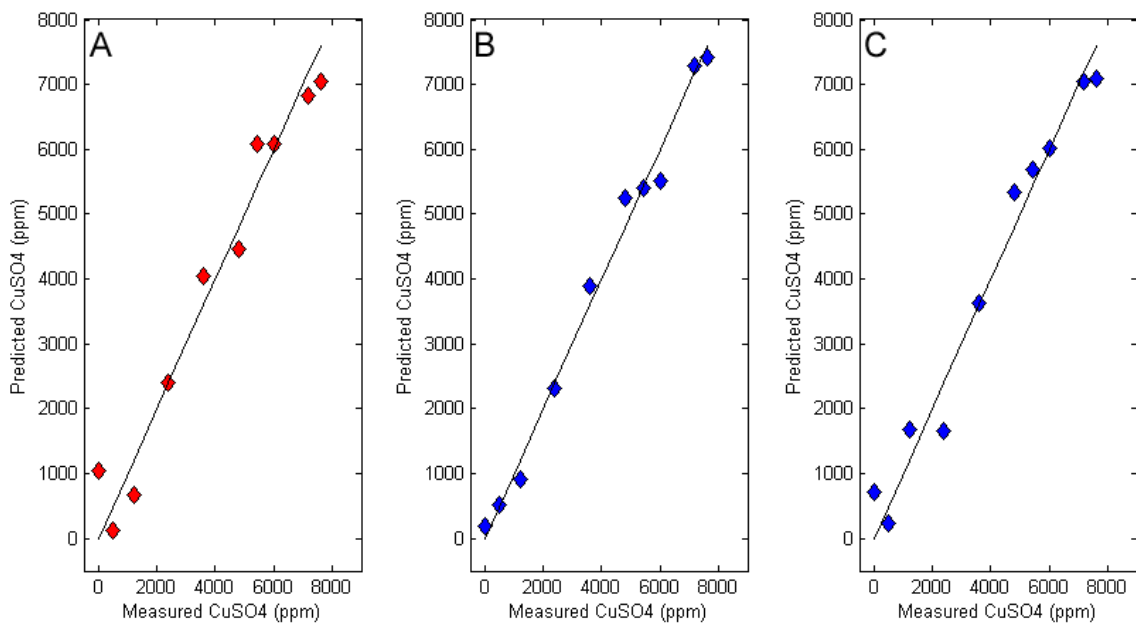


Figure 5.5: Leave-one-out cross validation plot of measured CuSO_4 versus predicted CuSO_4 for **A:** Hadamard data $r^2 = 0.989$, $\text{SEP} = 423$ ppm. **B:** CRISP resolution = 1, $r^2 = 0.997$, $\text{SEP} = 296$ ppm. **C:** CRISP resolution = 10, $r^2 = 0.992$, $\text{SEP} = 416$ ppm. Prediction errors were lower for CRISP scans indicating a clear advantage.

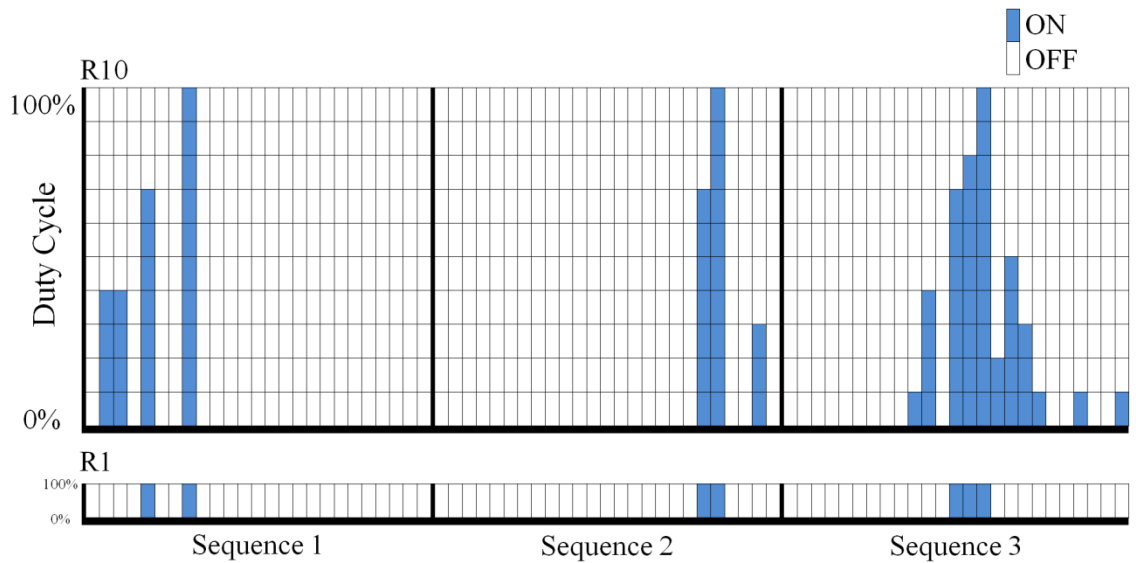


Figure 5.6: Positive pulse sequence duty cycles for CRISP scans at resolution equal to ten (R10) and one (R1). Each sequence corresponds to 25 LEDs. The value of a loading coefficient determines the number of pulses over the duty cycle and must be rounded to the nearest whole number corresponding to the scan resolution. PC loading vectors can be modeled more closely by increasing the scan resolution. For scan resolutions greater than 1, pulses are randomized over the duty cycle to produce as little correlation between LED pulse sequences as possible.

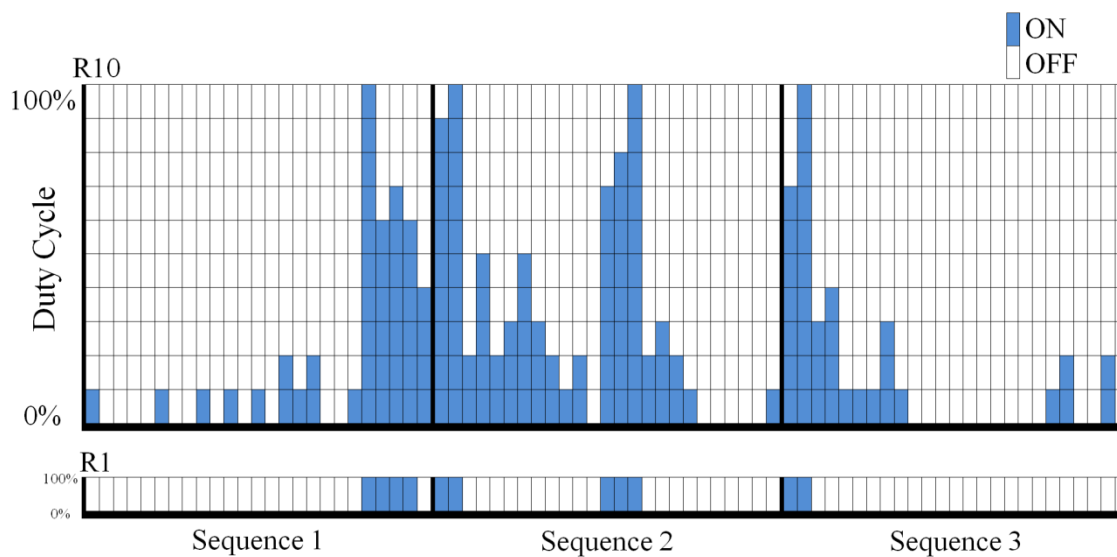


Figure 5.7: Negative pulse sequence duty cycles for CRISP scans at resolution equal to ten (R10) and one (R1). Each sequence corresponds to 25 LEDs. The value of a loading coefficient determines the number of pulses over the duty cycle and must be rounded to the nearest whole number corresponding to the scan resolution. PC loading vectors can be modeled more closely by increasing the scan resolution. For scan resolutions greater than 1, pulses are randomized over the duty cycle to produce as little correlation between LED pulse sequences as possible.

Chapter Six- BEST-CRISP for Classification of Common Martian Minerals and Identification of False Samples

Introduction

Modern analytical instrumentation increasingly employs more and more degrees of freedom in a single measurement process. The increased dimensionality of modern data sets makes them difficult to analyze and fosters a need for methods to extract useful information from the data. For example, Hyperspectral imaging (HSI) is capable of creating large volumes of data because images are acquired over many discrete wavelengths where the analyte is expected to have signals. Most images also contain a substantial number of pixels, and many images are often collected to form a video stream. HSI measures both physical and chemical properties. Data are represented as a cube of 3 or more dimensions, where two or three dimensions represent space and another dimension (or more) represents spectral intensities. HSI has been used throughout the space, pharmaceutical and defense industries. The Airborne Vis/IR imaging spectrometer (AVIRIS) was one of the first hyperspectral airborne systems launched in the late 1980s¹³⁷. AVIRIS collected imagery in 224 spectral bands over a range of 400 to 2500 nm. The use of hyperspectral sensors in satellite and aerial surveillance systems has motivated the streamlining of such units, reducing their form factor and increasing the efficiency and flexibility of the sensors. However, a significant amount of research has been and is still being undertaken in the development of efficient methods for interpretation of chemical signals and feature extraction¹³⁸.

Many post-data collection analysis methods have been developed for reducing high dimensional data into usable relationships between spectral features and physical properties. Principal component analysis (PCA) is a common tool used for preprocessing multivariate data, but PCA can be computationally expensive, requiring an $O(n^3)$ matrix factorization¹⁵. There exists a need for the development of real-time analysis techniques. Incorporating real-time methods that are computationally efficient are desirable for HSI and other spectroscopic methods to become more widely used. The development of rapid and efficient image analysis techniques, for example, will improve future HSI systems by eliminating post-collection data processing and limiting data storage needs. The reduction of HSI data to high-level information is an essential part of rapid analysis. One such method of reducing data to high-level useful information is by integrating the sensing and processing directly into the detector. ISP incorporates information about the sample directly into the physical design of the data collection process. A number of ISP techniques have been investigated for rapid quantification and qualification of analytes^{59, 107, 112}. ISP is accomplished by weighting an incoming signal so that the detector performs post-collection calculations. Performing calculations at the detector results in a signal that contains only high level information. This reduces the time spent for analysis because post-collection calculations are essentially done at the “speed of light”.

The Mars Science Laboratory (MSL) which is slated for a 2011 launch is expected to cost well over \$2 billion¹³⁹. The MSL mission is expected to generate a wealth of scientific data over its two year life. However at over \$2 billion the cost per analysis is very high. MSL has incorporated the ChemCam spectrometer which is capable of analyzing a target at a distance and will be used to guide the rover to “interesting”

locations where other onboard instruments can be used¹⁴⁰. The ChemCam is one step in maximizing the cost per analysis. However data generated by the ChemCam must be first analyzed before the rover can make the decision of whether to investigate further or not. In situations such as this where there is more than enough time for calibration before data collection but limited time after ISP can be especially useful. A rover utilizing ISP could make decisions in real-time thus maximizing the cost per analysis and minimizing time spent on calculations.

Astrobiology is the study of the origin, distribution and detection of life. Detection and identification of minerals such as hematites, calcites, feldspars, plagioclase and many others are of significant importance to astrobiologists¹⁴¹⁻¹⁴³. It is noted that geology plays a significant role in the survival of biological organisms¹²⁵. Endoliths are a type of extremophilic organisms that are capable of surviving in harsh environments because of their symbiotic relationship with the geologic environment. There is evidence that suggests Cyanobacteria organisms (which form phototrophic endolithic biofilms) existed earlier than 2.6 billion years ago and that these organisms may have played a role in converting the early atmosphere into an oxidizing one¹⁴⁴. If extant or extinct life exists on other solar bodies it is likely that it will resemble the extremophilic life found on Earth. UV/Visible to NIR reflectance spectroscopy is an ideal method for the detection of both inorganic and organic substances associated with extremeophiles. UV/Vis, NIR spectroscopy has been used for mineral identification, assessment of soil properties, biological and medical applications^{145-149, 77}. The use of UV/Vis to NIR wavelength spectroscopy combined with integrated sensing and processing would allow for the rapid identification of astrobiologically important materials.

This research describes an ISP algorithm which was developed for use with our solid state spectral imager (SSSI). The SSSI is a new device for acquiring hyperspectral images and spectroscopic data^{135-136, 150}. The current SSSI prototype comprises a 5x5 array of light emitting diodes (LEDs), single photodetector and on-board processor. The LEDs are sequenced according to the Hadamard transform¹⁵¹. Hadamard sequencing provides equal covariance and duty cycles on all diodes, reducing early diode failure. The Hadamard sequence progresses stepwise such that the decoded signal can be used to determine the contribution of each diode to the total spectrum. The SSSI contains no moving parts or lenses which facilitates the sizing down of the device. The current prototype constructed entirely from off-the-shelf parts is only 70 x 70 x 35mm in size (see Figure 6.1). Although the current prototype contains LEDs in the visible wavelength range, the SSSI can be used with diodes ranging from UV to NIR wavelengths.

CRISP scanning mode was developed for use with the SSSI. In CRISP scanning mode pulse sequences are varied over time to correspond to PC loading values. The loadings from PC analysis are selected to maximize the sample variation in a defined bandpass. The result of the integrated signal from a CRISP scan is a pseudo PC that directly corresponds to a specific analyte concentration or identity. The pseudo PC score is simply read as a voltage from the detector which can be accomplished with simplistic circuitry. The SSSI utilizing CRISP scanning mode has been shown to successfully quantify copper sulfate concentrations¹⁵².

ISP is essentially a pattern recognition algorithm. Problems arise in pattern recognition when the pattern-recognition algorithm encounters a sample unlike any in the original calibration set. This is termed the false sample problem. The false sample

problem is only an issue if the calibration set used to construct the ISP weights at the detector contained every substance in the known universe. An ISP routine is constructed from a training set of available materials which are expected to be encountered. For example an ISP routine for identifying Martian minerals would be constructed from minerals known to be found on Mars. It is possible however, that a new mineral composition or some other unexpected material might be encountered and identified as a training set sample. Calibration of an ISP algorithm would not have considered the new composition in the training set.

In order to deal with the false sample issues associated with the pattern recognition algorithm an ISP method that utilizes BEST distances has been developed. The BEST is a nonparametric method for estimating the distribution of a sample²⁵⁻²⁷. The BEST metric provides advantages over other methods because it does *not* assume equal covariance for all spectral groups and that each group is drawn from a normally distributed population. The BEST-CRISP algorithm calculates multidimensional standard deviations (MSDs) by converting the BEST model into a hash table. A training set is used to calculate a standardized set of distances and their centers which are stored. When a new sample is scanned its BEST distances is estimated by dividing the Euclidian distance of the stored center to the new sample by the Euclidian distance of the stored table value to the new sample. The on-line BEST-CRISP method is capable of estimating the BEST MSDs of a new substance at speeds orders of magnitude faster than the full BEST algorithm. The BEST-CRISP method allows for rapid determination of group classification and identification of false samples.

Theory

The motivation behind ISP is the reduction of complex data to high order information. BEST-CRISP begins with full spectrum analysis of a training set. Principal component analysis is employed to reduce the dimensionality of the training set data. In PCA a linear transformation of the data from one coordinate system to another of reduced dimensionality is performed. Calculation of PCs can be accomplished by performing the singular value decomposition (SVD) of a $m \times n$ matrix \mathbf{A} (the spectra) according to Equation 6.1.

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (6.1)$$

where \mathbf{U} ($m \times m$) is an orthonormal matrix representing PC scores, \mathbf{S} ($m \times n$) is a matrix of singular values, \mathbf{V}^T ($n \times n$) is an orthogonal matrix of eigenvalues (loadings). The dimensions of the matrices are dependent on the number of samples, m and the number of predictor variables, n . The decomposition of \mathbf{A} is calculated such that the maximum variation of the data set is found in the first PC. Each successive PC contains a lesser percent of the variation and is orthogonal to the previous one. Figure 6.2 illustrates the principal component scores of multidimensional spectra obtained from the USGS digital spectral library. In PCA spectra are transformed into spectral scores. Each signal, a recorded with a narrow bandpass is weighted at each wavelength, λ with a coefficient, f over n wavelengths.

$$\text{score} = f_1 a_{\lambda_1} + f_2 a_{\lambda_2} + f_3 a_{\lambda_3} + \dots + f_n a_{\lambda_n} \quad (6.2)$$

In PCA, the factors calculated by SVD are based on the linear variation in the spectral data set. In CRISP optical representations of the factors (f in Eq. 6.2) are created to

weight the signal (a in Eq. 6.2) by programming pulse sequences to match the loading vectors generated in PCA. LEDs can only occupy one of two states, on or off. In equation 6.2, the magnitude of the loading coefficient f_i , signifies the contribution of the i^{th} wavelength (LED) to the *score*. The on and off states of a CRISP sequence with varying duty cycles must be integrated over time so that the summed value corresponds to a factor score. In this implementation of CRISP, more pulses are used per integration frame at wavelengths with higher loading values (see Fig. 6.3). The CRISP modulation bit rate (bandwidth) controls the intensity resolution of the factor result. CRISP pulse sequences are randomized to produce as little covariance between the sequences as possible. The length of sequences used for each loading depends on the desired resolution. At the lowest resolution, a CRISP sequence would encode all loading values as a binary one or zero. To generate the binary sequences the loading values, f in Eq. 6.2 are first normalized according to Equation 6.3.

$$f_{norm} = f_i / f_{max} \quad (6.3)$$

Each normalized loading value corresponding to an LED is multiplied by the desired resolution. The resolution simply describes the number of pulses that will be used per an integration sequence. If for example, the desired loading resolution is 100 and f_{norm} at $i = 1$ in Eq. 6.3 is equal to 0.5 then the LED corresponding to $i = 1$ will have a 50% duty cycle over 100 pulses. As the number of pulses and integration frames is increased, the intensity resolution can become quite high, allowing the loading coefficients from PCA to be closely simulated. The binary representations are randomized for each diode duty cycle to produce as little covariance as possible between the duty cycles.

In CRISP mode the emitted light resembles the desired loading spectrum. Each loading vector contains positive and negative values. Because negative intensities are impossible the loadings are divided into positive and negative vectors. Therefore for each PC used in classification there are two CRISP pulse sequences (one + and one -). The CRISP pulse sequences can be used to classify a new spectrum. A second training set is used to generate pseudo PCs (voltage scores). The voltage scores are obtained from the summed signal over each CRISP pulse sequence according to equation 6.4.

$$V = \sum_j^J f_j a_{\lambda_j} \quad (6.4)$$

where f_j is the loading at wavelength j , a_{ij} is the spectral amplitude of the i^{th} mixture at wavelength j , V is the detector voltage and j is an index over J frequencies. In this situation V corresponds to a PC and is directly proportional to sample identity. Negative pulse sequences are inverted and summed with the positive counterpart to generate a single voltage score. Just as PC scores of same samples cluster together in hyperspace the voltage scores will also cluster together.

To determine the class of each new spectrum a bootstrap distribution is created for each group in the training set of voltage scores. Each bootstrap replicate is standardized to the calculated BEST MSD according to Equations 6.5.

$$D = \frac{d(C,B)}{BESTSD} \quad (6.5)$$

Where $d(C,B)$ is the Euclidian distance from the center C , $(1 \times n)$ of the bootstrap distribution to the bootstrap point, B $(m \times n)$, $BESTSD$ $(m \times n)$ is a vector of BEST standard deviations for each B to C , and D $(m \times 1)$ is vector of standardized distances for

each bootstrap point \mathbf{B} . The dimensions of the matrices are dependent on m the number of bootstrap replicates and n the number of PCs used to classify the samples. The BEST method is a nonparametric classification technique. BEST draws probability density contours around each sample population. Each normalized bootstrap, \mathbf{B}_N ($m \times n$) is calculated according to Equation 6.6.

$$\mathbf{B}_N = \mathbf{C} + [D * (\mathbf{B} - \mathbf{C})] / \|\mathbf{B} - \mathbf{C}\| \quad (6.6)$$

The normalized bootstraps, \mathbf{B}_N are stored in a hashtable along with the centers of each distribution. For each new CRISP scan, the resulting voltage \mathbf{V} is centered to a distribution and normalized according to Equation 6.7.

$$\mathbf{V}_N = \mathbf{V} - \mathbf{C} / \|\mathbf{V} - \mathbf{C}\| \quad (6.7)$$

The hash table is searched for the bootstrap point \mathbf{B}_N , closest to the test sample \mathbf{V}_N by calculating the dot product of \mathbf{V}_N and \mathbf{B}_N . The largest dot product of \mathbf{B}_N and \mathbf{V}_N indicates the bootstrap, \mathbf{B}_N , is closest to the test sample \mathbf{V} . The BEST SD of \mathbf{V} is estimated according to Equation 8.

$$BEST\ SD = \frac{d(\mathbf{C}, \mathbf{V})}{d(\mathbf{B}, \mathbf{V})} \quad (6.8)$$

Where $d(\mathbf{B}, \mathbf{V})$ and $d(\mathbf{C}, \mathbf{V})$ are Euclidian distances.

Materials and Methods

Materials. Mineral spectra used in simulations were obtained from the USGS digital spectral library¹⁵³. The following spectra were used: calcite, hematite, olivine, bytownite, and orthoclase. Real mineral samples were obtained from the University of Kentucky

geology department. Samples were selected to closely match the description in the USGS database as much as possible. Three false samples were used for identification. The false samples were seashell, moss and wet hematite. The false samples were chosen to demonstrate components that would be of considerable interest if found but that were not included in a training set.

Simulated Data. Simulations were performed in order to determine how the modified BEST algorithm performed compared with the original, and the efficiency gains afforded by the modification. Simulated mineral spectra were replicated and random noise was added to produce spectra with signal-to-noise ratios with respect to peak height between 20 and 45. Simulated false sample spectra were constructed by randomly sorting the sample spectra at each wavelength. For m samples at n wavelengths the intensity values of n were randomly sorted on m . This was done so that the false sample spectra would contain the same spectral variation between samples at each wavelength.

Data Collection. Five replicate scans of each mineral sample were acquired in random order. Each replicate was scanned 45 times, rotating the sample three times. This produced 225 spectra for each sample. The scans were averaged to give 5 spectra for each sample. The data was scatter corrected and blank scans were subtracted. All samples were scanned in ambient light. The prototype SSSI comprised 25 LEDs with five each of the following wavelengths: 470, 565, 595, 609, and 635 nm. A fiber optic probe was attached to the SSSI and positioned approximately 1cm from the sample target. The target area was approximately 1 cm^2 . The SSSI was controlled via a serial connection to a laptop computer. A graphical user interface (GUI) was created in MatLab 7.0.1 (The Mathworks Company, Natick, MA, USA) to send commands to the

SSSI CPU and transfer all data. Data was imported into MatLab for analysis. PCA was performed on all data. The PCs that showed the best separation between groups were used for construction of the CRISP codes.

CRISP Sequence Construction. Only three PCs were used to classify the samples; therefore, their corresponding loadings were used to generate the CRISP codes. For each loading the positive and negative loading intensities were separated. The absolute value of the negative loading data values were used because optical intensities cannot be negative. Each loading vector was normalized separately according to Eq. 6.3. The length of a CRISP code is determined by the number of bits of intensity resolution used. The CRISP code resolution was set to 10; therefore, for each LED (wavelength) the maximum number of pulses used to emulate a loading value was 10. Figure 6.4 indicates the duty cycle for each diode over the pulse period.

Results and Discussion

Simulation. In BEST-CRISP, BEST MSDs are estimated by comparing the voltage output from a CRISP scan to a hash table of standardized distances. The BEST MSD computation using a hash table significantly reduces the computational requirement because the bootstrap distribution and BEST MSDs of each bootstrap are pre-computed.

As a comparison running BEST-CRISP in 3 dimensions with 5 groups (150 samples per group) and a hash table comprising 10,000 values (5 groups x 2000 bootstrap points each) required less than 2 seconds to perform. The full cross-validated BEST computation for all 5 groups took over 230 seconds. The calculations were done on a laptop computer with a 1.58GHz processor and 2GB of RAM. The intra- and inter-cluster distances for the simulated data were compared to determine how well the BEST-

CRISP method modeled BEST MSDs. Figure 6.5 is a comparison of the inter-cluster distances for the full BEST algorithm and the BEST-CRISP method. The average inter-cluster distances for the BEST-CRISP and the full BEST methods were 50.21 and 50.71, respectively. Average intra-cluster distances were 0.79 and 0.80 for the BEST-CRISP and BEST methods, respectively.

The BEST-CRISP method performed comparably with the full BEST method. The sensitivity and specificity for classification with each method were both 100 and 97.5%, respectively. Sensitivity and specificity were calculated as:

$$Sensitivity = \frac{TP}{(TP+FN)}$$

$$Specificity = \frac{TN}{(TN+FP)}$$

Where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative.

A simulated scan was performed with false sample spectra included to determine the ability of the BEST-CRISP method to identify samples not included in the original CRISP training set. In total 1500 spectra were simulated (750 true samples + 750 false samples) and their BEST MSDs were approximated using a hash table comprising 10,000 values. The BEST-CRISP method successfully identified all false samples with sensitivity and specificity of 100%.

Hadamard Scans. Data from actual mineral samples was acquired with the SSSI. Three PCs containing 97.6% of the variation were used to separate the training set samples. Figure 6.6 illustrates the variation among the sample groups. The BEST

method was used to classify the data. The sensitivity and specificity for classification was 100 and 99%, respectively. The SSSI prototype comprised 5, wavelengths; therefore the mineral samples were selected to maximize the variations of the SSSI wavelengths. Figure 6.7 indicates the between group separation achieved with the SSSI at 5 wavelengths. CRISP codes were generated for the mineral samples using the PC loadings corresponding to the top three PCs. The resulting CRISP sequences comprised three positive followed by three negative pulse sequences. The sequences were constructed to have scanning resolution of 10. Therefore the duty cycle for each SSSI diode varied over six sequences consisting of 10 pulses each.

CRISP Scans. Voltage data from mineral scans were acquired with the SSSI in CRISP scanning mode. Because CRISP utilizes pulse sequences that are weighted functions of the training set data the resulting intensity signals should correspond to a specific sample in the training set. The sensitivity and specificity for the classification of minerals using the full BEST algorithm was 100 and 95%, respectively. The specificity was slightly worse for CRISP scans compared to Hadamard scans because hematite was classified as olivine indicating a false positive for olivine. This can be attributed to the small number of training samples (5 samples) and wavelengths used. Minerals are not uniform in nature, and it is expected that a larger number of training samples and excitation energies would be needed to maximize the variations that exist between samples. However, all samples were correctly identified. Figure 6.8 illustrates the BEST MSDs between groups for CRISP scans. Inspection of Figures 6.7 and 6.8 illustrates the difference between full wavelength (Hadamard) and CRISP scans. The average inter-cluster distances for Hadamard and CRISP scans were 12.9 and 27.1, respectively.

Average intra-cluster distances for Hadamard and CRISP scans were 1.25 and 1.01, respectively. The intra-cluster distances are minimized, and the inter-cluster distances are maximized with CRISP scanning indicating a clear advantage over Hadamard scanning. The CRISP scan data were used to construct a hash table to be used in subsequence BEST-CRISP analysis.

BEST-CRISP Scans. The purpose of the BEST-CRISP hashing metric is to identify when a false sample situation is likely to exist. BEST MSDs are estimated by comparing each intensity value from a CRISP scan to a hash table of precomputed distances. Figure 6.9 illustrates the BEST-CRISP calculated MSDs between groups from BEST-CRISP scans. BEST MSDs are calculated on-line using the hash method. The sensitivity and specificity for the classification of minerals using the BEST-CRISP method was 96 and 93%, respectively. The specificity of the method to classify samples decreased because the intra-cluster distances increased yielding a higher number of false positives. The average inter-cluster distances were greater for both the CRISP and BEST-CRISP methods (27.1 and 18.55, respectively) indicating a clear advantage of the methods over Hadamard scanning (average inter-cluster distance of 12.9). The Hadamard mode resulted in an average intra-cluster distances of 1.25 while CRISP and BEST-CRISP resulted in a distance of 1.01 and 1.74 respectively. Because MSDs are estimated from a set table when variations in test samples increase from the training set it is more difficult to determine class. It is not surprising that non-uniform samples such as minerals would be difficult to classify this way. A larger training set encompassing a larger amount of variation can be used to increase the accuracy of the method. However, the lower sensitivity calculated with this method would seem as a disadvantage, but

because BEST-CRISP is very sensitive to variations in the sample it is very sensitive to false samples.

To determine the ability of the BEST-CRISP metric to identify false samples three new sample sets were introduced. In total 15 false sample scans were acquired with the true minerals. Figure 6.10 indicates the separation achieved between the mineral and false samples. Table 6.1 indicates the accuracy, precision and sensitivity for identifying false samples from actual minerals. Accuracy and precision were calculated as follows:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

$$Precision = \frac{TP}{(TP+FP)}$$

All false samples were correctly identified yielding sensitivity of the method at 100%. Specificity was calculated to be 96% because some actual mineral samples were classified as false samples. However, no false sample was ever identified as belonging to the training set.

Conclusion

Bootstrap error estimated single sample technique complementary randomized integrated sensing and processing has been explored as a real-time analysis for the classification of mineral samples and the identification of false samples. A SSSI with preloaded codes could be used to identify specific analytes in remote environments. The SSSI is a low cost and rugged spectrometer. The SSSI has potential uses in remote environments such as Mars where it could be deployed on a rover to help guide to areas of interest. Night time measurements are also possible with the SSSI because it is not dependent on

ambient lighting. In addition, fluorescence imaging at night can also be performed with appropriate UV emitting diodes.

Chapter Six Tables

Table 6.1: BEST-CRISP analysis of 15 false samples included into the mineral sample set. Samples were classified to any group within three standard deviations in hyperspace. Samples with SDs greater than 3 to any group were considered false samples.

Group	Correct Classification of False sample	Accuracy (%)	Precision (%)	Sensitivity (%)
Orthoclase	15	100	100	100
Calcite	15	100	100	100
Hematite	15	95	93.75	100
Bytownite	15	100	100	100
Olivine	15	100	100	100
Average	15	99	98.75	100

Chapter Six Figures

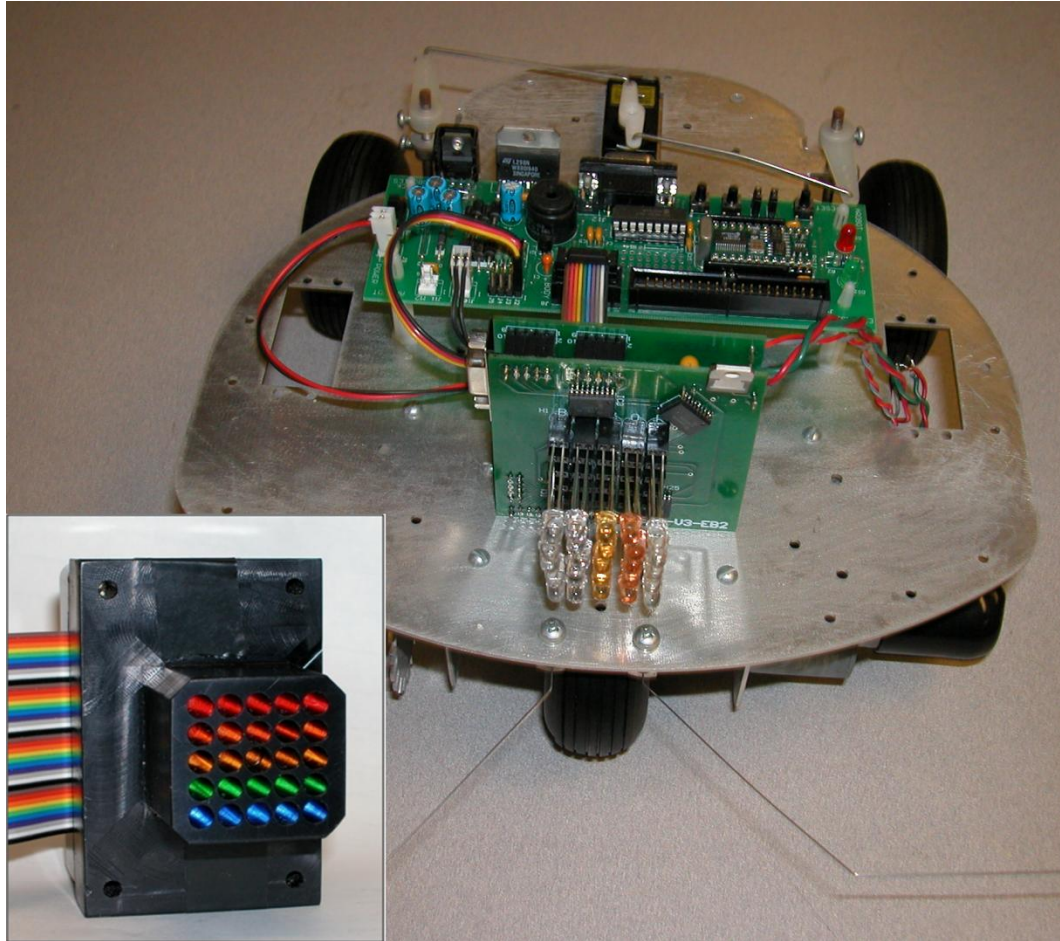


Figure 6.1: The small size of the SSSI allow it to be easily mounted to a robot. The current SSSI prototype comprises a 5x5 LED array in the visible wavelength region and measures 70x70x35mm. The SSSI diodes must be collimated (inset) to for increased spatial resolution. The free roving robot can be directed to interesting locations indicated by the SSSI.

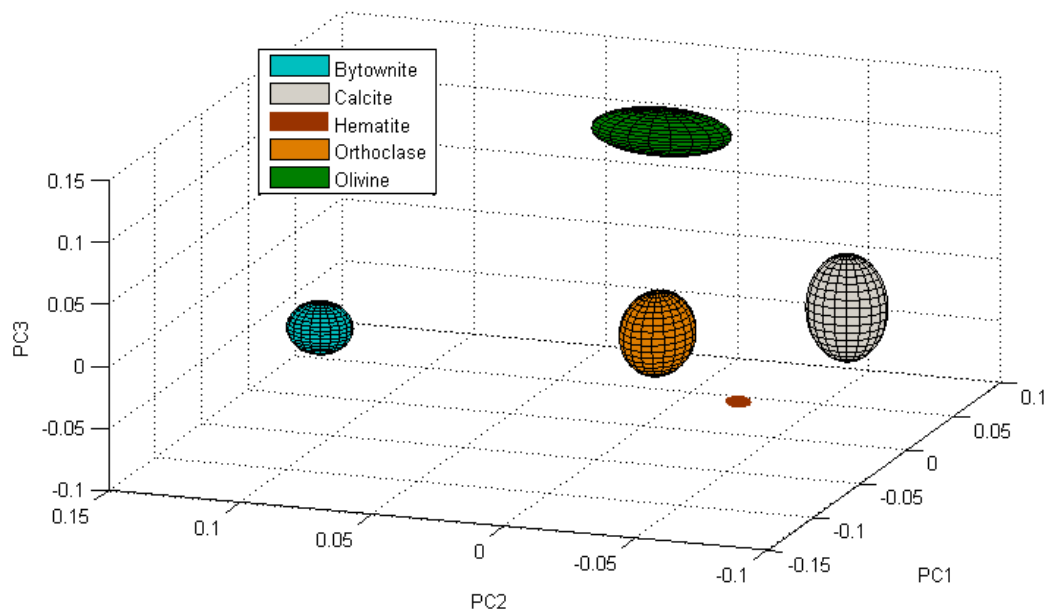


Figure 6.2: The PCs that captured the largest variation between groups were plotted against each other. The ellipsoids represent a 3SD contour for each mineral type. Each ellipse contains 150 points with a S/N ratio of 20 – 45.

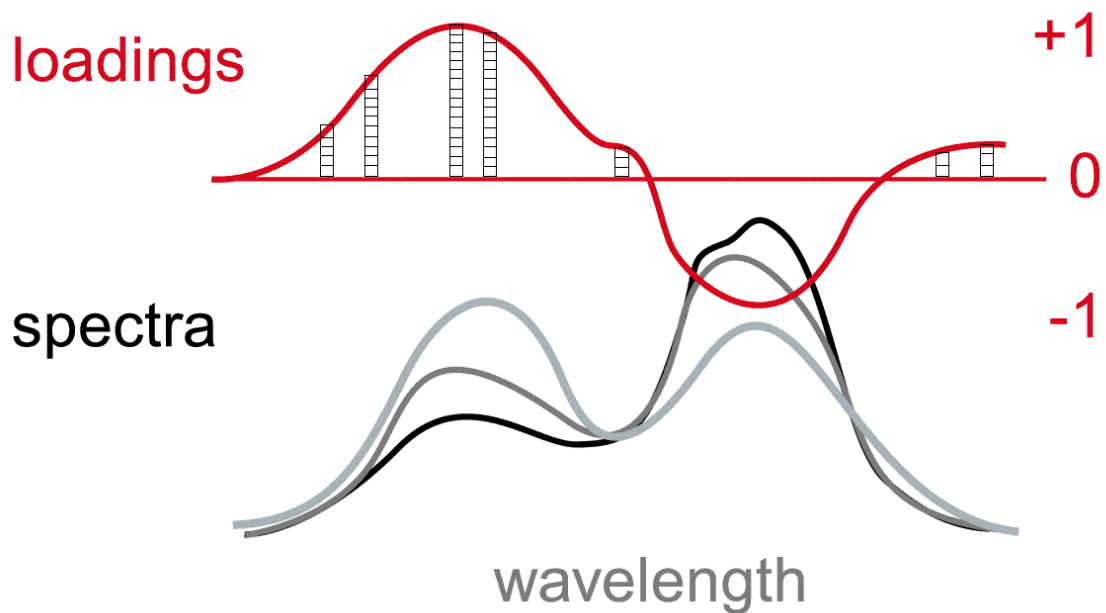


Figure 6.3: CRISP is employed by varying the duty cycle of the pulses (pulses represented by blocks in loadings graph) at each diode wavelength, so that in each integration frame, the most pulses are delivered at the wavelengths where the most light is needed to make the integration sum to the analyte value sought.

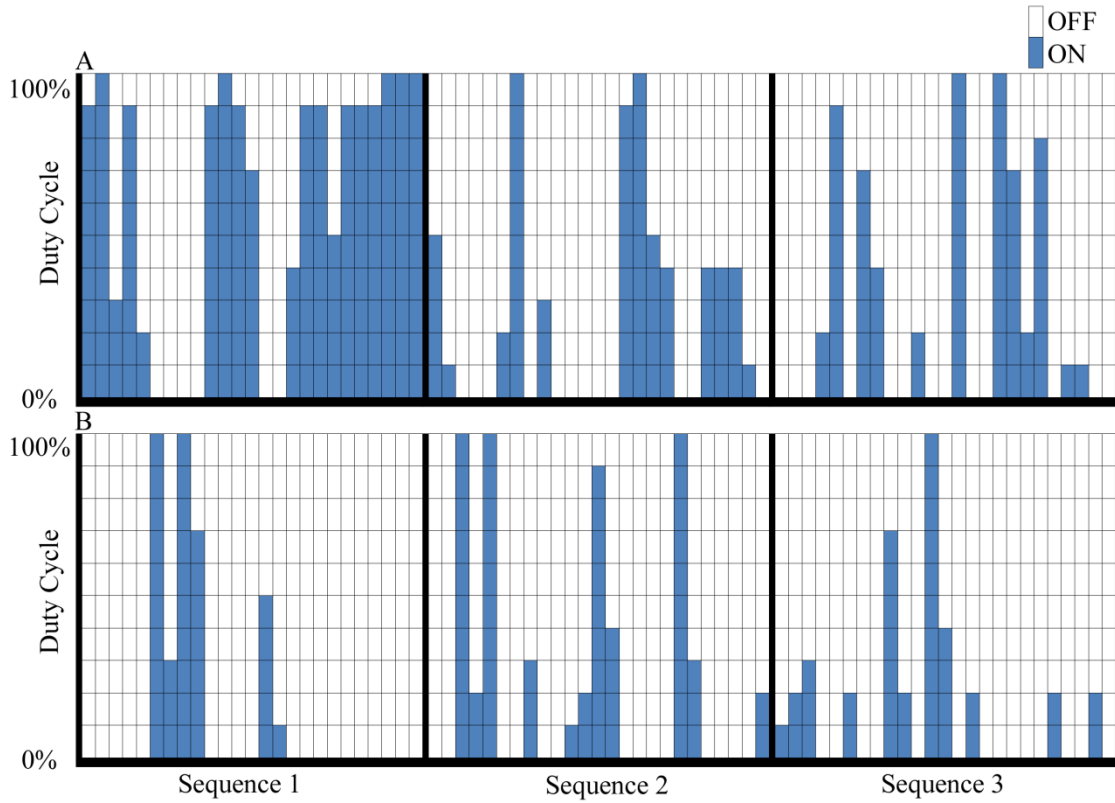


Figure 6.4: Positive (A) and negative (B) loading pulse sequences for a CRISP code with resolution of 10. Pulses are randomized over the duty cycle to produce as little correlation between LED pulse sequences as possible. Each loading sequence corresponds to 25 LEDs.

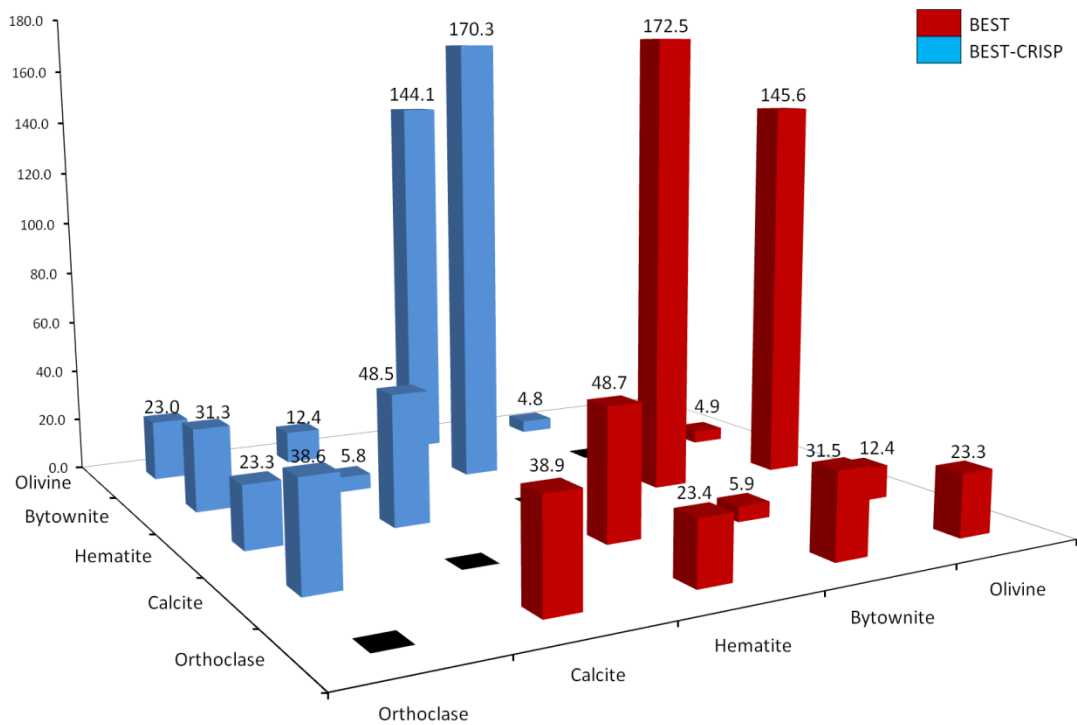


Figure 6.5: Comparison of BEST-CRISP to the full BEST algorithm calculated MSDs for inter-cluster distances. Average inter-cluster distances between groups were 50.21 and 50.71 MSDs for BEST-CRISP and BEST, respectively. Average intra-cluster distances (not shown on the diagonal) for the BEST-CRISP and BEST methods were 0.79 and 0.80, respectively.

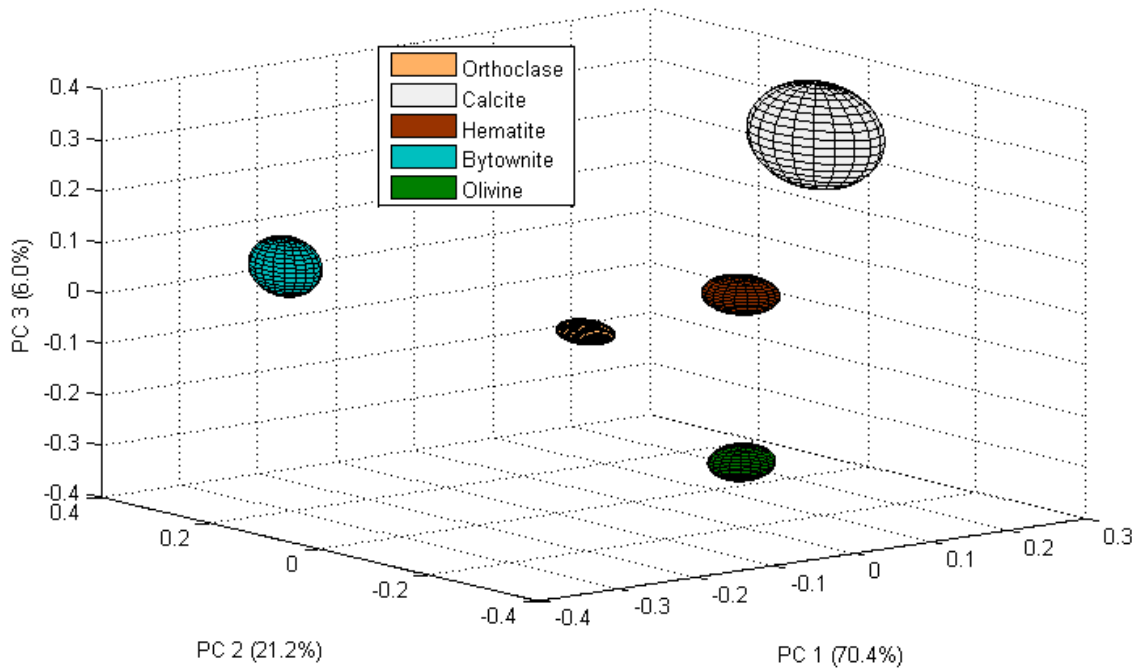


Figure 6.6: The variation between mineral groups can be viewed by projection of the top three PCs into three-dimensional space. The ellipses indicate a one standard deviation contour level for each group.

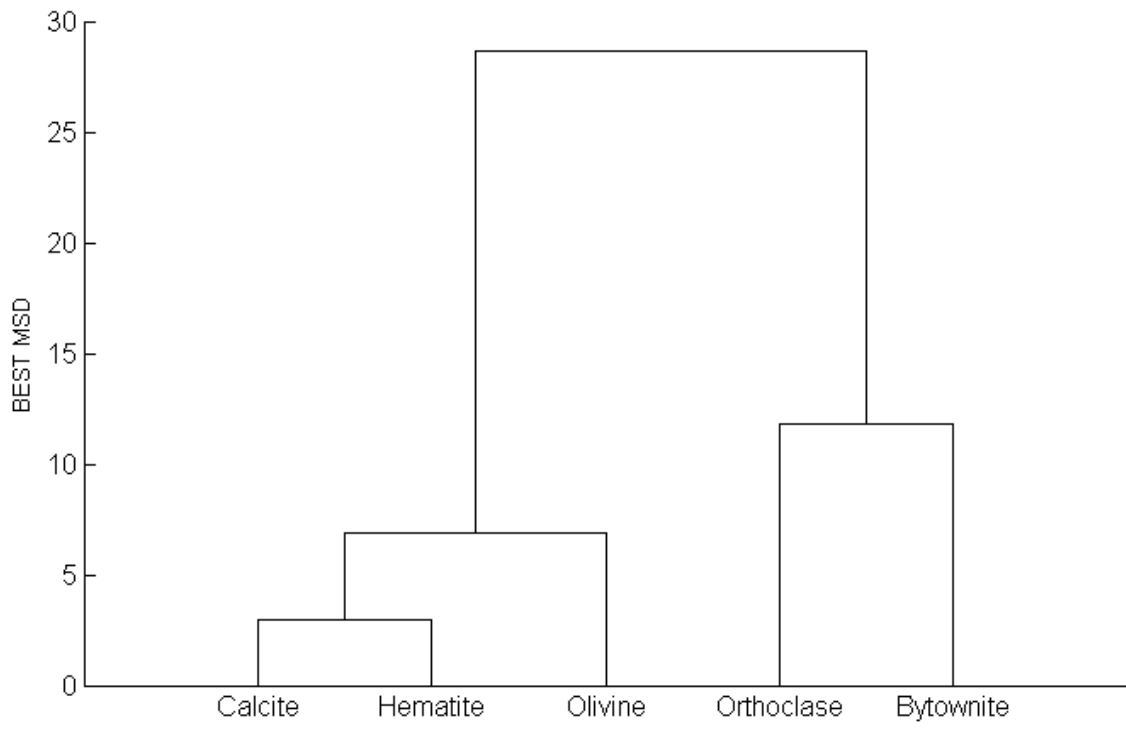


Figure 6.7: Dendrogram plot indicating BEST MSDs between groups for Hadamard scan data. Groups are linked according to their largest BEST MSD.

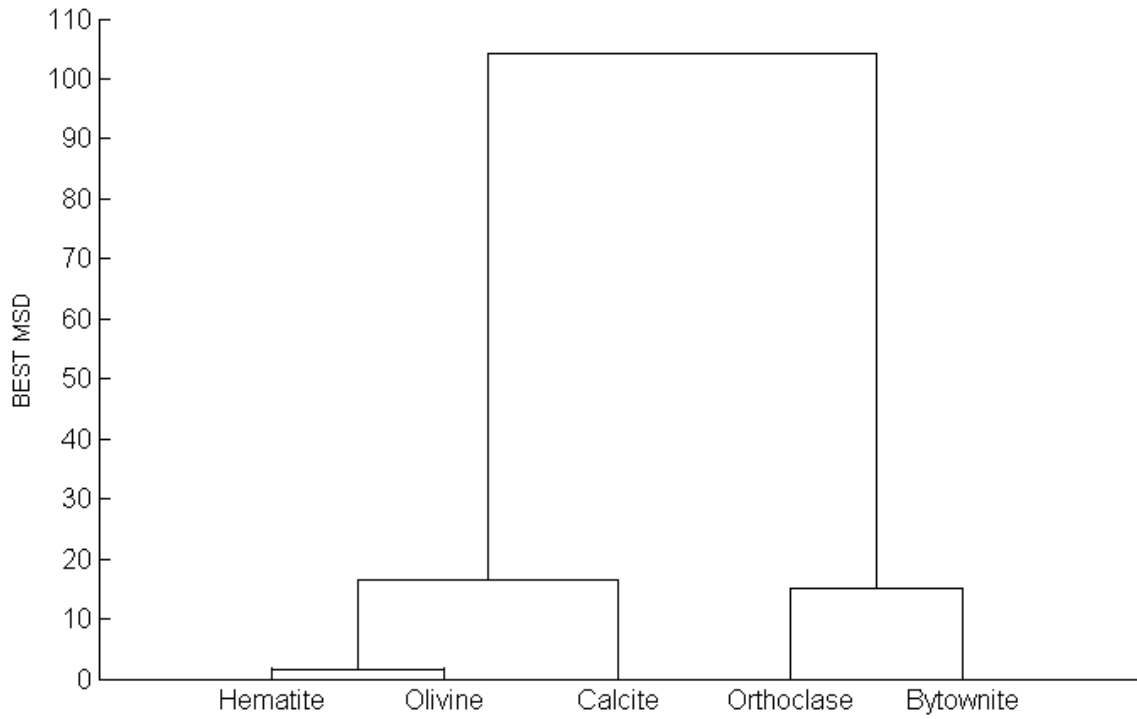


Figure 6.8: Dendrogram plot indicating BEST MSDs between groups for CRISP scan data. Groups are linked according to their largest BEST MSD.

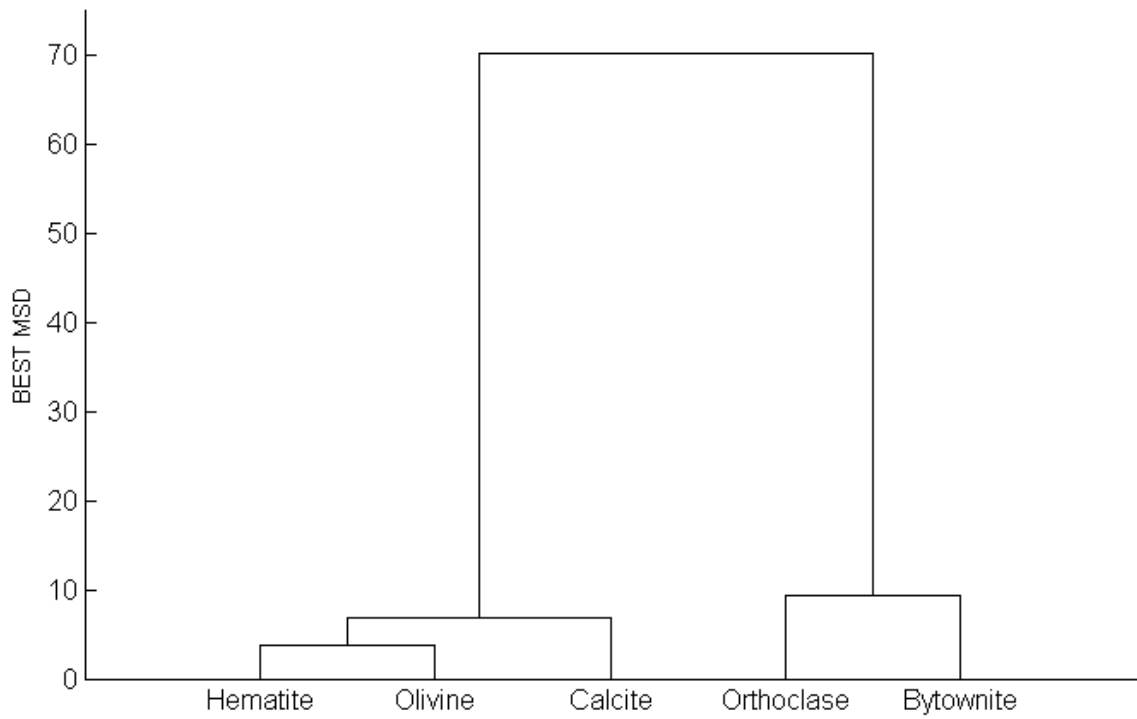


Figure 6.9: Dendrogram plot indicating BEST MSDs between groups for BEST-CRISP scan data. BEST MSDs are calculated on-line. Groups are linked according to their largest BEST MSD.

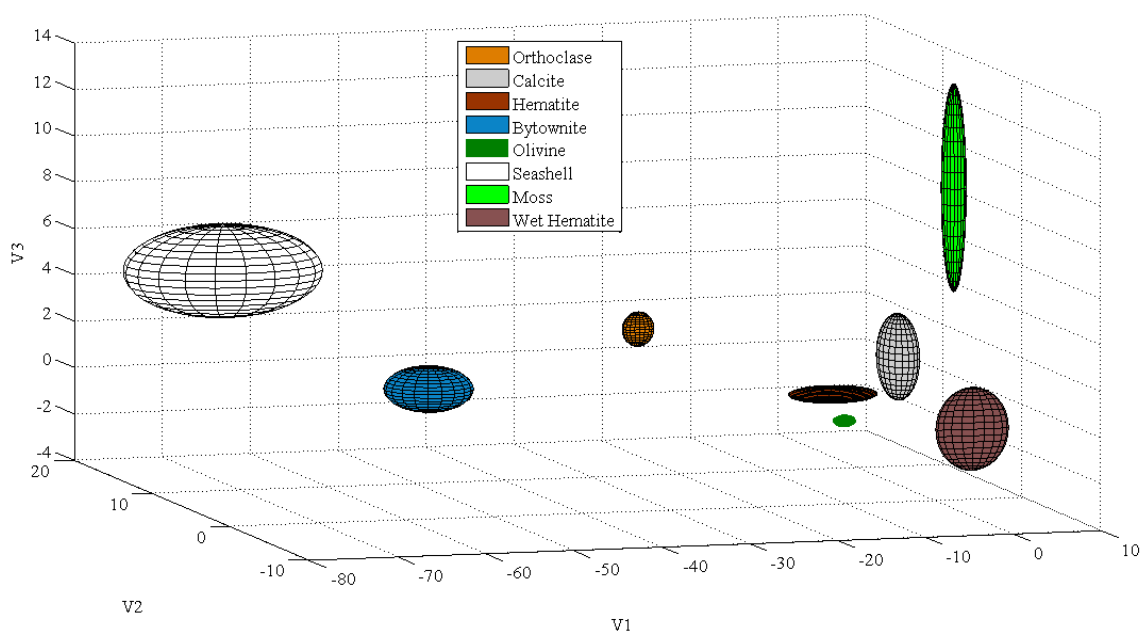


Figure 6.10: The variation between mineral groups and false samples can be viewed by projection of the acquired voltage into three-dimensional space. The ellipses indicate a one standard deviation contour level for each group.

Thaddaeus Hannel, Robert A. Lodder. In preparation for submission to *Analytical Chemistry*. 2009

Chapter Seven- Size Does Matter: Solid State Spectral Imager (SSSI) on a Chip and the Search for Extraterrestrial Life

Introduction

It is likely that no human will ever physically explore the outer reaches of the Galaxy, barring the actual invention of some sort of warp or wormhole technology like that which exists in science fiction¹⁵⁵. It would take an explorer traveling at the speed of light over 95,000 years just to traverse the Milky Way. While we may never physically venture out of our Galaxy we may however, extend ourselves into the far reaches of space through technology. Any science mission that attempts to explore other solar systems will no doubt search for life. It is not the purpose of this paper to debate the notion that life may exist elsewhere in the universe; however, if we consider that there are on the order of 10^{21} stars in the universe then the statistical probability that Earth is the only life-containing planet is likely low. The Drake equation is often used to estimate the probability of finding extraterrestrial life.

Drake equation. The Drake equation was named after its creator Dr. Frank Drake, who conceived it in 1961. The equation

$$N = R * f_p n_e f_l f_i f_c L \quad (7.1)$$

where, N is the number of communicative civilizations, R is the rate of formation of Sun-like stars, f_p is the fraction of stars with planets, n_e is the number of Earth-like worlds per planetary system, f_l is the fraction of Earth-like planets where life develops, f_i represents the fraction of planets where intelligence develops, f_c is the fraction of planets in which

Earth-like communications technology develops, and L represents the lifetime of communicating civilizations, estimates the number of advanced communicative civilizations in a Galaxy. Current estimates using the Drake equation suggest that approximately 10,000 communicative civilizations may exist in the Milky Way¹⁵⁶. Unfortunately, there has been no way yet to prove the Drake equation. However, the possibility of life on other planets in the Galaxy cannot be ignored. Furthermore, if Drake's equation estimates so many communicative civilizations in a single galaxy, then the number of life forms without the ability to communicate is probably much higher.

Contact. It has been hypothesized that if intelligent civilizations exist, then they too would search for life elsewhere in the universe^{154,157}. However, the vastness of space and the technological barriers that complicate exploration of it may equate the search for intelligent life to finding a needle in a cosmic haystack. Any contact with a more advanced civilization may occur long after that civilization has become extinct. For example, in 76,000 years the Voyager probes will emerge in the Alpha Centauri star system. Any contact with an advanced civilization, extant or extinct, would have an astounding impact on humanity. However, until contact is established the search for any form of life will be ongoing. If humanity can establish the presence of life elsewhere in the solar system, then perhaps the Fermi question¹⁵⁸ can be partially put to rest.

The Barrier. Special relativity indicates that a mass cannot travel at speeds that meet or exceed that of light. This limitation arises because the mass of the object increases to infinity as the speed of the object approaches the speed of light. Simply put, the larger the mass the more energy is required to move it. Therefore physics benefits the

small when it comes to the energy required to move an object through space, and the smaller the object is the faster it can be moved given the energy available.

The closest star system to the Sun is Alpha Centauri (approximately 4.2 light years)¹⁵⁹. If a ship could travel at one third the speed of light a trip to Alpha Centauri would only take 12.6 years. The amount of energy required to propel a space craft at one third the speed of light is tremendous, however, if the space craft were sufficiently small enough it might be feasible to do so with current and future technologies^{160,161}.

Micro- and Nanotechnologies

Micro- and nanoelectronics are technologies in which the control or manipulation of electronics is accomplished on the microscale (100 – 0.1 μm) and nanoscale (100 -1 nm), respectively. The advantages of operating in these small-scales are in the efficiency (the amount of power dissipated in a circuit is proportional to its size) and speed (propagation delay is reduced as size is reduced) of the electronic circuitry. While there are still many technological barriers that must be overcome before these technologies are totally realized, many advancements in this area are taking place. Carbon nanotubes for example, have been used as AM demodulators, RF detectors and RF mixers^{162,163}.

Solid State Spectral Imager. The Solid-State Spectral Imager (SSSI) is a new instrument for gathering hyperspectral images and spectroscopic data of biological and geological samples. Using a laser-diode array, photodetector, and on-board processing, the SSSI combines innovative spectroscopic integrated sensing and processing with a hyperspace data-analysis algorithm. Ultraviolet (UV), visible, and near- infrared laser (NIR) diodes in the current SSSI illuminate target points using a precomputed sequence,

and a photodetector records the amount of reflected light. For each point illuminated, the resulting reflectance data are processed to separate the contribution of each wavelength of light and classify the substances present. The SSSI has the ability to operate for extended periods of time and in extreme environments.

The SSSI can be used to search for extremophilic organisms. Extremophilic organisms are organisms that live and thrive in extreme conditions; commonly identified as poor-oxygen, extreme heat and extreme cold. Such organisms are found in various regions on Earth such as the Antarctic, Atacama Desert; Chile, and deep ocean heat vents to name a few¹⁶⁴. In order to identify such compounds common to extremophilic organisms, laser diodes with wavelengths ranging from the UV to NIR could be used¹³⁵.

The current SSSI prototype comprises a 5x5 light emitting diode (LED) array, a photodetector circuit, and a microcontroller. A PC running MatLab utilizes a serial connection to send scan commands to the unit. The SSSI can operate in normal scanning or complementary randomized integrated sensing and processing (CRISP) scanning mode. In normal scanning operation the LEDs are pulsed in an orthogonal pseudorandom (Hadamard) sequence. The orthogonal pseudorandom sequence progresses step wise such that the demultiplexed signal can be used to determine the contribution of each LED wavelength. The entire sequence takes less than 5 ms. Using normal scanning mode would require each scanned sample to be analyzed and compared to a database which would require transmitting large amounts of data back to Earth or the ability of the device to determine the identification of the analyte itself. In CRISP scanning mode the duty cycle of the orthogonal pseudorandom code is varied for different target analytes to produce a situation in which the integrated signal at the detector is directly proportional

to the concentration of one specific analyte. In this manner the energy and time needed to process the data are eliminated and the analog signal at the detector corresponds directly to the analyte concentration. CRISP codes are compiled into a table on Earth using the Hadamard sequence to perform calibration by scanning substances that might be encountered on other planets, such as biological or geological specimens. Once the spectral data are acquired, principal component analysis (PCA) or a similar factor analytic method are used to examine the spectral variances among the data set. PCA is a multivariate analysis technique used commonly for data reduction in spectroscopic analysis^{57,167}. The PCs are generated from the wavelength data such that the first PC contains most of the variation of the original data set. The PCs that correspond to the most significant spectral variances are used to create the CRISP codes. The loadings associated with the PCs are used to indicate the wavelength regions that show the greatest variance (see Fig. 7.1). These wavelengths are the most important in creating a prediction equation for concentrations that is linear. More pulses are used per integration frame at wavelengths with higher loadings (see Fig. 7.2). The CRISP scanning sequence incorporates the chemometric analysis into the transducer, which significantly reduces the processing time and power needed to make a unique identification. All that is needed is to classify the output voltage of the detector, which can be accomplished using anything from a voltmeter (one PC) to the Bootstrap Error-adjusted Single-sample Technique (BEST, for multiple PCs). (The BEST method of sample classification calculates the distance between data clusters in multidimensional standard deviations²⁷.)

Instead of using the Hadamard scanning mode, a SSSI loaded with CRISP codes could be used to detect life forms by their chemistry, shape, and motion on far away

planets. Water is simple to detect using its strong near-infrared absorption spectrum as the basis for a CRISP code. If CRISP sequences are used (which identify expected compounds using a library) and unexpected response is recorded, the Hadamard scanning mode can be reactivated to collect raw spectral data and transmit the information back to Earth for human analysis.

SSSI Downsizing. Unlike other spectrometers the SSSI has no moving parts or lenses, so it can be reduced in size very easily. The SSSI consists of three main parts that could be integrated onto a single chip: microcontroller, laser diodes with electronic switches, and detector. The current SSSI prototype is assembled with COTS (commercial off the shelf) parts and weighs approximately 70 g. Most of the mass of the prototype comprises the printed circuit board and connections. It should be possible with existing technology to create a SSSI such that all of the components are integrated with microelectronic products¹⁶⁸. At present time the microcontroller chip currently measures 16 x 16 x 1.2 mm; integration of the SSSI to a single chip would greatly reduce the mass and the size of the device.

Communicating with the SSSI

Typical communication systems operating at X- and Ka-band frequencies require large power sources, receivers and transmitters. The transmitting and receiving antennas require a significant amount of space because their optimum size increases with wavelength. Antenna size is not a substantial concern when the space craft is itself very large; however, power consumption is always a concern in space. One advantage of microwave frequencies is the ability to propagate through most dielectric materials,

which can be a disadvantage when using higher frequencies associated with optics. However, optical frequencies comprise a much larger bandwidth. To meet the bandwidth and size requirement of a microprobe, high telecommunication frequencies will need to be used. Optical (Infrared, Visible and UV) and terahertz (Far-IR) frequencies offer the high bandwidth and gain in a small antenna needed for deep space communications.

Optical Communications. Optical frequencies have significant advantages over the typically used X- and Ka-bands, as optical communication systems are shown to have lower mass and DC power usage¹⁶⁹. Optical frequencies however, generally require line of sight and suffer from scattering effects by dust and particulates in the atmosphere. In order to utilize optical communications systems for deep space missions it may be necessary to employ an Earth orbiting monitoring system which would transmit optical data received to Earth via traditional radio frequencies. Optical communications used onboard a small probe sent into another planetary system may have difficulty sending transmissions as the signal-to-noise (s/n) ratio would be lowered by the system's star (unless the transmission distances are shortened by deploying a network of small probes for communication). Near-infrared laser wavelengths will likely be more useful than visible laser wavelengths because there is less scattering and absorption of the longer wavelengths as they cross the galaxy.

Terahertz Communications. Terahertz (THz) energy is considered to be at the boundary between optical and electric sources because it lies in the region near the far-IR. THz wavelengths are in the micrometer range and thus do not require a large antenna

size and also do not suffer from atmospheric scattering due to dust and other particulates. However, THz frequencies do suffer from attenuation of atmospheric water vapor¹⁷⁰.

Companies like Virginia Diode already offer solid-state devices for generating THz light. Furthermore, optical rectification and the electro-optic effect are reciprocal processes in a nonlinear optical crystal, so a single small crystal can potentially serve as an electro-optic transceiver to alternately transmit pulsed electromagnetic radiation (optical rectification) and detect the return signal (electro-optic effect) in the same crystal. Such devices might be manufactured easily for use as communication devices in micro and nanoscale space probes¹⁷¹. Under shock some crystals also emit terahertz light¹⁷², and may some serve as small sources of electromagnetic waves.

Optical and THz communications strategies still require significant technological developments before they can be used on a mission to another planetary system. Due to problems with atmospheric scattering either strategy alone may not work. Optical and THz strategies may provide a synergistic solution to communications aboard interstellar missions and have potential due to their small size and lower power consumption.

NASA and Microtechnology

NASA's New Millennium Program (NMP) was begun in 1995 to speed the development of new technologies. A few of the missions flown and envisioned under the NMP entailed the use of small satellites and probes. Among the smallest probes used were those of Deep Space 2 in which two probes weighing less than 1.2 kg were sent to Mars.

While the probes were completely passive and ultimately failed in their mission, they illustrate a growing ability to use small-scale technology in planetary missions¹⁷³.

Physically small probes will likely also be small in power reserves. Networks of minute probes communicating digitally with light pulses may have to be set up to enable galactic communication. Where networks are not practical, the small size of the devices may permit them to be moved through space, carrying their data onboard until they reach their destination and are able to download their data. For all of these reasons, the spectroscopic and chemical analysis ability of single-chip SSSIs using near-IR laser communication networks may make them useful as probes for extraterrestrial life.

Conclusion

It is clear that due to our physical limitations humans may never venture beyond Earth. While further developments in micro- and nanotechnologies are still needed, we are entering an era in which the exploration of the galaxy beyond our own solar system is looking increasingly possible.

Chapter Seven Figures

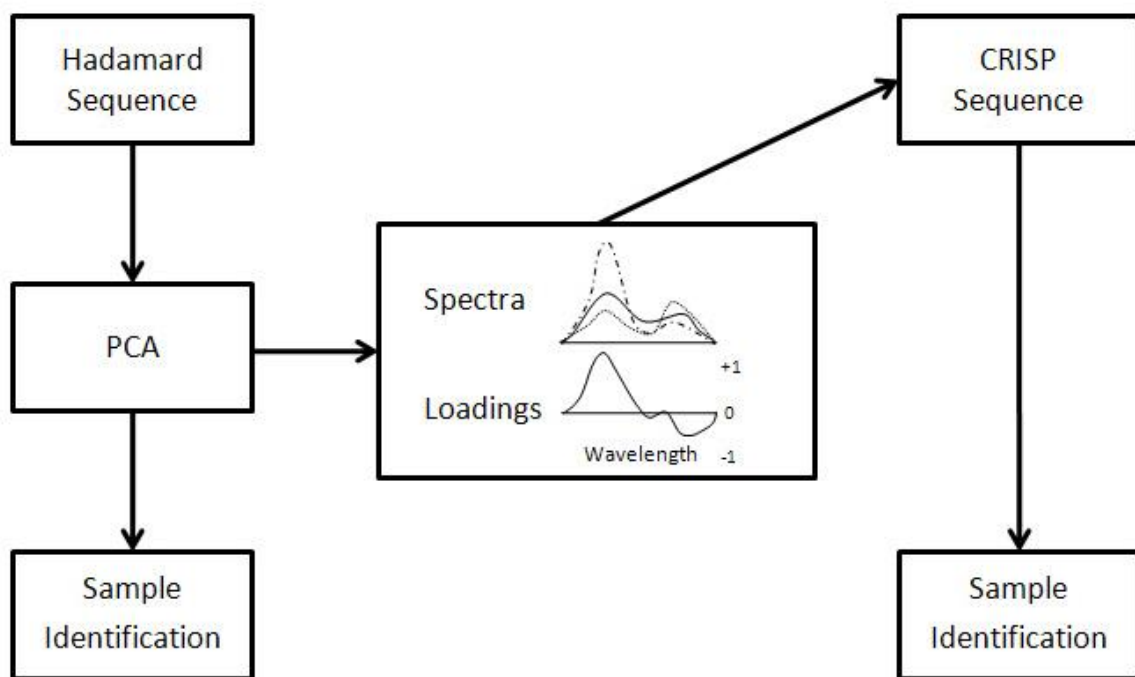


Figure 7.1. The Hadamard sequence is used to obtain the spectral information from a sample set. PCA is used to determine the wavelength regions that show the greatest sample variation. The CRISP sequence is generated from the wavelength regions that show the greatest variance among the sample set. In Hadamard scanning mode chemical quantification requires that each spectrum be analyzed by a computer. In CRISP scanning mode, the quantification of the sample constituents is instantaneous because the analysis has been integrated into the detector.

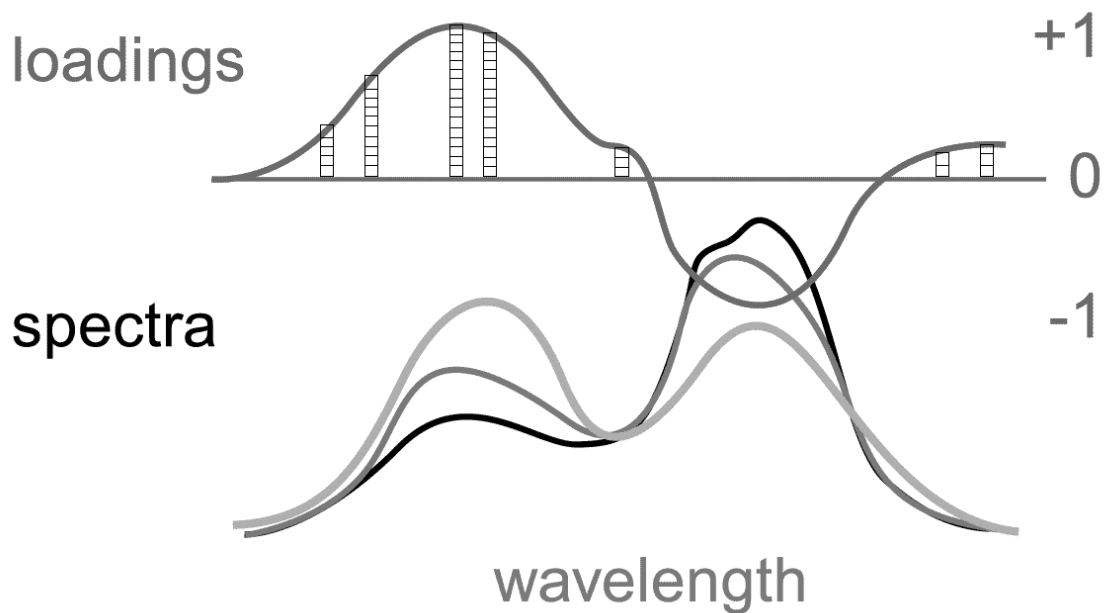


Figure 7.2. CRISP is employed by varying the duty cycle of the pulses (pulses represented by blocks in loadings graph) at each diode wavelength, so that in each integration frame, the most pulses are delivered at the wavelengths where the most light is needed to make the integration sum to the analyte value sought

Section V: Conclusion of Dissertation

The work contained in this dissertation demonstrates the clear advantage of PRISMs for data and processing reduction in sensing systems through implementation of chemometrics into the physical design of the detector. The demand for PRISMs is motivated by different goals. In the pharmaceutical industry the move from cGMP to PAT calls for alternative techniques that are cost effective, rapid, and nondestructive. The ability of the manufacturing process to be self-guided by real-time measurements can lead to better use of materials and increased savings in both time and cost. To this end, the potential application of SCARS for differentiating formulations of a first-in-class diabetes drug was demonstrated. BEST-SCARS successfully identify false samples with sensitivity and specificity of 92.0 and 100%, respectively. Near-IR HSI also has potential for use in the pharmaceutical and biotech industries. The problems with existing alcohol monitoring treatments make a noninvasive and surreptitious system of interest. MFC near-IR HSI was demonstrated as a method for noninvasive monitoring of BAC. BEST-MFC NIR HSI successfully indicated facial regions which corresponded to erroneous BAC values with sensitivity and specificity of 99.8 and 100%, respectively. A near-IR imaging system comprising solid MFC filters is a robust and inexpensive solution to traditional HSI methods.

In astrobiology small, inexpensive, and robust sensors are needed for space and terrestrial applications. Exploration of terrestrial environments is one such area where conventional spectroscopic methods suffer in terms of efficiency. To address this problem CRISP-SSSI was investigated for the quantification of copper sulfate solutions

and identification of mineral samples. BEST-CRISP successfully identified false samples with sensitivity and specificity of 100 and 96.0%, respectively.

In this dissertation PRISMs were applied to both acoustic and optical spectroscopy. The application of PRISMs indicated a clear advantage of time, cost, and efficiency over the existing traditional techniques. The development of a BEST-ISP algorithm makes the use of PRISMs possible. In every case the BEST-ISP method was shown to detect false samples with *sensitivity* > 90% and *specificity* > 90% on categorical data.

Appendix A: MatLab Code

A.1 BEST-ISP Code

```
function [Hashtable, cnter] = GenHash(data,numBreplicates,ngroups)

%A Bootstrap distribution with size equal to numBreplicates is created for each group
(ngroups) %in data. The Hashtable is created from the normalized distances to each
bootstrap replicate %from the center (Cnter) of the distribution.

% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

[nrows ncols] = size(data);
incr = nrows/ngroups;
cnter = zeros([ngroups ncols]);

cnt = 1;
cnt2 = 1;
for i = 1:incr:nrows
    [NBdist(cnt:cnt+(numBreplicates-1),:),cnter(cnt2,:)] = NormBestDist(data(i:i+(incr-
1),:), numBreplicates,1);
    [Hashtable(cnt:cnt+(numBreplicates-1),:)] =
SortDistance(NBdist(cnt:cnt+(numBreplicates-1),:), cnter(cnt2,:), numBreplicates);
    cnt = cnt+numBreplicates;
    cnt2 = cnt2+1;
end
```

```

function [NBdist,cnter] = NormBestDist(train, numreps,SDs)

%Generates nomalized BEST distances from the training set train. numreps
%is the desired size of the bootstrap. SDs is the standard deviation of the
%normalized distances. This option is only for plotting NBdist contours.
%SDs should always be set to 1. cnter is the center of NBdist.

% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

[r c] = size(train);
[btrain,cnter]=replica(train,numreps);

eucdis = zeros([1,numreps]);
for i =1:numreps
eucdis(i) = sqrt(sum((cnter-btrain(i,:)).^2));
end

sd = zeros([1,numreps]);
sds = zeros([1,numreps]);
sdskew = zeros([1,numreps]);
for i = 1:numreps
[sd(i),sds(i),sdskew(i)] = qb(train,btrain,btrain(i,:),cnter,1,0);
end

Distances = (eucdis./sdskew);

NBdist = zeros(size(btrain,1),c);
for i = 1:numreps
NBdist(i,:) = cnter+ (btrain(i,:)-cnter)*(Distances(:,i)*SDs)/norm(btrain(i,:)-cnter);
end

```

```

function [Sdist] = SortDistance(NBdist, cnter, numBreplicates)

% This function sorts a set of normalized BEST distances NBdist with center, %cnter.
% The size of NBdist for a single group in NBdist is numBreplicates. A smaller
% set of the sorted distances Sdist can be selected by taking every 5, 50,
% or 100 distance for example. Because the distances
% are sorted, when taking a specified number of the distances an equal
% distribution of the distances will be selected. This is done when the full
% distribution is not required.

% :: Copyright 2009 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

[nrows ncols] = size(NBdist);
nDist = zeros(nrows,ncols);

for i = 1:numBreplicates
nDist(i,:) = NBdist(i,:) - cnter;
end

theta = zeros(1,nrows);

for i = 1:numBreplicates
theta(i) = acosd((dot(nDist(1,:),nDist(i,:)))/(norm(nDist(1,:))*norm(nDist(i,:))));
end
theta = theta';
[s ind] = sort(theta);
Sdist = NBdist(ind,:);

```

```

function [BEST] = BEST_Hashloop(newspec,cnter,hashtable,numgroups)

%This function uses a set of stored normalized BEST distances (hashtable)
%to estimate the BEST SD of the test data newspec. The input cnter is a table of
%centers from the hashtable distributions for each group. The number of
%groups are defined by the input numgroups. The output BEST contains the estimated
%BEST SDs.

% :: Copyright 2009 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

l = size(hashtable,1)/numgroups;

for i = 1:numgroups
[pick(i)] = hashdotprod(newspec,cnter(i,:),hashtable(i*l - (l-1):l*i,:));
temphash = hashtable(i*l - (l-1): l*i,:);
[BEST(i)] = Bestcalc(temphash, pick(i), cnter(i,:), newspec);
end

```



```

function [pick] = hashdotprod(newspec,cnter,hashtable)

%This function centers the newspec to the cnter of the normalized bootstrap
%distributions in hashtable. newspec is normalized (standardized)
%according to the center of the bootstrap distribution with which it is being compared.

% :: Copyright 2008 :: Thad Hannel, Aaron Urbas & Dr. Robert Lodder, ASRG,
University of
% Kentucky

newspec = newspec - cnter;
%unit vector
normspec = newspec/norm(newspec);

%For the operation dot A=B so multiplication is used instead.
% when u and v are column vectors dot(uv) = u'*v
% When u and v are row vectors dot(uv) = u*v'
dotprods = (hashtable*normspec');

%The maximum dot product indicates the bootstrap that is in the direction
%of the sample spectrum.
pick = find(dotprods == max(dotprods));

%Because the sample spec has been normalized to the center of the bootstrap
%distribution the closest bootstrap point will now have the maximum inner space
%between the point. Pick selects this point from the HashTable
pick = pick(1);

```

```

function [BEST] = Bestcalc(Vdist, pick, cnter, testspec)
%The BEST SD is estimated by calculating by the following procedure.
%Calculate the Euclidian distance between the closest Bootstrap replicate (Vdist),
%defined by pick, and the test spectrum, testspec d(B,X). Calculate the Euclidian
%distance between the center of the bootstrap distribution, cnter and the
%test spectrum, testspec d(C,X). The BEST SD can be estimated by d(C,X)/d(B,X).

% :: Copyright 2009 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

BX = pdist([Vdist(pick,:);cnter]);
CX = sqrt(sum((cnter(1,:)-testspec).^2));
BEST = CX./BX;

```

```

function [CVstat] = CVstatsBEST(BESTSDs,ids,FStot)

%Claculaes sensitivity and specificity for identifying false samples.
%BESTSDs are BEST SDs calculated from the Hashtable output from BEST_Hashloop
%ids are according to true samples. FStot is the total number of false sample spectra.

% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
% Kentucky

% Calculates accuracy, precision and recall in percent according to:
%
% Accuracy = (TP + TN) / (TP + TN + FP + FN)
% Precision = TP / (TP + FP)
% Recall = TP / (TP + FN)

if min(ids) == 0
    ids = ids+1;
end

uids = unique(ids);
nids = length(uids);
[r c] = size(ids);

% accuracy = zeros(nids,1);
% precision = accuracy;
% recall = accuracy;
% tcorrect = accuracy;
% sensitivity = accuracy;
% specificity = accuracy;

% correct = zeros(r,1);
% for i = 1:r
%     correct(i) = classes(i,ids(i));
% end
%TN = True sample identified as true BESTSD < 3
%FP = True sample identified as false BESTSD > 3
correctA = BESTSDs(1:r,:);
correctA(correctA<3)=1;
correctA(correctA>3)=0;

%TP = False sample correctly identified: BESTSD >3
%FN = False Sample not identified: BESTSD <3
correctB = BESTSDs(r+1:end,:);
correctB(correctB<3)=0;
correctB(correctB>3)=1;

```

```

for i = 1:nids
    tot = length(ids(ids==uids(i)));
    TP = sum(correctB(:,i));
    FN = FStot - TP;
    TN = sum(correctA(ids==uids(i),i));
    FP = tot - TN;

    accuracy(i,:) = ((TP+TN)/(TP+TN+FP+FN))*100;
    precision(i,:) = (TP/(TP+FP))*100;
    recall(i,:) = (TP/(TP+FN))*100;
    tcorrect(i,:) = TP;
    sensitivity(i,:) = (TP/(TP+FN))*100;
    specificity(i,:) = (TN/(FP+TN))*100;
    tp(:,i) = TP;
    tn(:,i) = TN;
    fp(:,i) = FP;
    fn(:,i) = FN;
end

```

```

CVstat.accuracy = accuracy';
CVstat.precision = precision';
CVstat.recall = recall';
CVstat.correct = tcorrect';
CVstat.specificity = specificity';
CVstat.sensitivity = sensitivity';
CVstat.sensT = sum(tp)/(sum(tp)+sum(fn));
CVstat.specT = sum(tn)/(sum(fp)+sum(tn));

```

A.2 SSSI MatLab Code

```
function [CLP,CLN,CRISPP, CRISPn] = generateCRISP(loads, resolution)
%Generates CRISP sequences from PCA loading data
%loadings correspond to the PCs used in cross validation
%resolution sets the number of pulses for each CRISP sequences
% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
Kentucky

[r c] = size(loads);
LP = loads;
LN = loads;
LP(LP<0)=0;
LN(LN>0)=0;
LN = -LN;
CLP = zeros([r c]);
CLN = zeros([r c]);
for i = 1:r
    if max(LP(i,:))==0
        CLP(i,:) = LP(i,:);
    end
    if max(LP(i,:))~=0
        CLP(i,:) = (LP(i,:)./max(LP(i,:)))*resolution;
    end
    if max(LN(i,:))==0
        CLN(i,:) = LN(i,:);
    end
    if max(LN(i,:))~=0
        CLN(i,:) = (LN(i,:)./max(LN(i,:)))*resolution;
    end
end

CLN = round(CLN);
CLP = round(CLP);
CLP = reshape(CLP',1,r*c);
CLN = reshape(CLN',1,r*c);

outp = zeros(max(CLP),length(CLP));
outn = zeros(max(CLN),length(CLN));
for j = 1:length(CLP)
    outp(1: CLP (j),j) = 1;
    outn(1: CLN (j),j) = 1;
end
```

```
%Randomize on/off states
[r2 c2] = size(outp);
for i = 1:c2
    random = rand(resolution,1);
    [r ind] = sort(random);
    outpR(:,i) = outp(ind,i);
    outnR(:,i) = outn(ind,i);
end
CRISPP = outpR;
CRISPN = outnR;
```

```

function [Sig] = simulateCRISP(CRISPP, CRISPN, Data, numisp)

% This function simulates a CRISP scan on Data
% CRISPP and CRISPN are generated from generateCRISP.m
% numisp = numloadings used to create the crisp codes
% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
Kentucky

[r c] = size(CRISPP);
[rd cd] = size(Data);
l = size(CRISPP,2)/numisp;
cnt = 1;
for k = 1:l:c % for wavelengths over the number of loadings
    for j = 1:r % resolution of the CRISP code
        for i = 1:rd % number of samples
            SigP(i,j,cnt) = dot(CRISPP(j,k:(l-1)+k),Data(i,:));
            SigN(i,j,cnt) = dot(CRISPN(j,k:(l-1)+k),Data(i,:));
        end
    end
    cnt=cnt+1;
end

SigP = sum(SigP,2);
SigP = squeeze(SigP);
SigN = sum(SigN,2);
SigN = squeeze(SigN);
Sig = SigP - SigN;

```

```

function [Hexoutp, Hexoutn] = genhex(CRISPP, CRISPN)

%This function generates the 2 bit hex code for the SSSI processor
% CRISPP and CRISPN are generated from generateCRISP.m
%Hexoutp and Hexoutn are the hex code corresponding to the positive and negative
loading
%data, respectively.
% :: Copyright 2008 :: Thad Hannel & Dr. Robert Lodder, ASRG, University of
Kentucky

[r c] = size(CRISPP);
Seqs = c/25;

for i = 1:Seqs
Cp((i*5)-4:i*5,:) = reshape(CRISPP(:,(i*25)-24:i*25)',[5 r*5]);
end

for i = 1:Seqs
Cn((i*5)-4:i*5,:) = reshape(CRISPN(:,(i*25)-24:i*25)',[5 r*5]);
end

for i = 1:Seqs
Cdatap((i*(r*5))-((r*5)-1):i*(r*5),:) = Cp((i*5)-4:i*5, :)';
end

for i = 1:Seqs
Cdatan((i*(r*5))-((r*5)-1):i*(r*5),:) = Cn((i*5)-4:i*5, :)';
end
Cdatanum = num2str(Cdatan);
Cdatapum = num2str(Cdatap);

Hexoutp = dec2hex(bin2dec(Cdatapum),2);
Hexoutn = dec2hex(bin2dec(Cdatanum),2);

```


SSSI_GUI_Hadamard.ini File

BAUDRATE= 115200

HADAMARDLEDS= 31

HADAMARDSEQS= 31

SATURATION= 16777216

XVALS= 0.8 0.9 1 1.1 1.2 1.8 1.9 2 2.1 2.2 2.8 2.9 3 3.1
3.2 3.8 3.9 4 4.1 4.2 4.8 4.9 5 5.1 5.2

SEQS=

```
1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0
0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1
1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0
0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0
1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1
0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
1 0 1 0 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
0 1 1 1 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1 1 1 0 0 1 1 0 0 0 0 1 1
1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0
0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0
1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1
0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0
1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1
0 1 1 0 0 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1
1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0
0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
1 0 1 0 1 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1
0 1 1 0 0 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1
1 1 0 0 1 1 0 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 0 1 1 0 0 1 1 0
0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
```

1 0 1 1 0 1 0 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 1 0 1 1 0 1 0
0 1 1 1 1 0 0 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 0 0 1 1 1 1 0 0
1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 1 0 1 1 0 0 1 1 0 1 0 0 1

SSSI_GUI_CRISP.ini File

BAUDRATE= 115200

CRISPLEDS= 25

CRISPSEQS= 6

SATURATION= 16777216

XVALS= 1 2 3 4 5 6

SEQS=

% SEQS are user defined. See Appendix B for details.

```

function varargout = SSSI2_H_C_GUI(varargin)

% This GUI is used to run the SSSI in Hadamard or CRISP
Modes

%       SSSI2_H_C_GUI, by itself, creates a new
SSSI2_H_C_GUI or raises the existing
%       singleton*.
%
%       H = SSSI2_H_C_GUI returns the handle to a new
SSSI2_H_C_GUI or the handle to
%       the existing singleton*.
%
%
SSSI2_H_C_GUI('CALLBACK', hObject, eventData, handles,...)
calls the local
%       function named CALLBACK in SSSI2_H_C_GUI.M with the
given input
%       arguments.
%
%       SSSI2_H_C_GUI('Property','Value',...) creates a new
SSSI2_H_C_GUI or raises the
%       existing singleton*. Starting from the left,
property value pairs are
%       applied to the GUI before
SSSI2_H_C_GUI_OpeningFunction gets called. An
%       unrecognized property name or invalid value makes
property application
%       stop. All inputs are passed to
SSSI2_H_C_GUI_OpeningFcn via varargin.
%
%       *See GUI Options on GUIDE's Tools menu. Choose "GUI
allows only one
%       instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help
SSSI2_H_C_GUI

% Last Modified by GUIDE v2.5 17-Aug-2008 03:21:34

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',  gui_Singleton, ...

```

```

        'gui_OpeningFcn',
@SSSI2_H_C_GUI_OpeningFcn, ...
        'gui_OutputFcn',
@SSSI2_H_C_GUI_OutputFcn, ...
        'gui_LayoutFcn', [] , ...
        'gui_Callback', []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State,
varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
end
% End initialization code - DO NOT EDIT

% --- Executes just before SSSI2_H_C_GUI is made visible.
function SSSI2_H_C_GUI_OpeningFcn(hObject, eventdata,
handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)
% varargin   command line arguments to SSSI2_H_C_GUI (see
VARARGIN)

% Choose default command line output for SSSI2_H_C_GUI
handles.output = hObject;

% initial Flags for GUI
handles.SSSI_open=0;
handles.choice=0;
handles.cr_scount=1;
handles.hd_scount=1;
handles.cr_scan(1,60)=0; % needs to be more versatile in
future
handles.hd_scan(1,25)=0; % needs to be more versatile in
future

set(handles.reset, 'Enable', 'off');
set(handles.com_port, 'Enable', 'off');

```

```

set(handles.com_open, 'Enable', 'off');
set(handles.bar_graph, 'Visible', 'off');
set(handles.s_no, 'Enable', 'off');
set(handles.s_no_v, 'Enable', 'off');
set(handles.model, 'Enable', 'off');
set(handles.model_v, 'Enable', 'off');
set(handles.firmware, 'Enable', 'off');
set(handles.firmware_v, 'Enable', 'off');
set(handles.num_leds, 'Enable', 'off');
set(handles.num_leds_v, 'Enable', 'off');
set(handles.num_seq, 'Enable', 'off');
set(handles.num_seq_v, 'Enable', 'off');

reset_gui(hObject, eventdata, handles);
% Update handles structure
guidata(hObject, handles);

% UIWAIT makes SSSI2_H_C_GUI wait for user response (see
UIRESUME)
% uiwait(handles.figure1);
end

% --- Outputs from this function are returned to the
command line.
function varargout = SSSI2_H_C_GUI_OutputFcn(hObject,
eventdata, handles) %#ok<INUSL>
% varargout cell array for returning output args (see
VARARGOUT);
% hObject handle to figure
% eventdata reserved - to be defined in a future version
of MATLAB
% handles structure with handles and user data (see
GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;
end

%-----%
-----%
%-----Object creation
functions-----%
%-----%
-----%

% --- Executes during object creation, after setting all
properties.

```

```

function experiment_choice_CreateFcn(hObject, eventdata,
handles) %#ok<INUSD>
% hObject    handle to experiment_choice (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    empty - handles not created until after all
CreateFcns called

% Hint: popupmenu controls usually have a white background
on Windows.
%         See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

end

%-----
%-----%
%-----Object Call Back
functions-----%
%-----
%-----%

% --- Executes on selection change in experiment_choice.
function experiment_choice_Callback(hObject, eventdata,
handles)
% hObject    handle to experiment_choice (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

% Hints: contents = get(hObject,'String') returns
experiment_choice contents as cell array
%         contents{get(hObject,'Value')} returns selected
item from experiment_choice
end

% --- Executes on button press in load_config_button.
function load_config_button_Callback(hObject, eventdata,
handles)
% hObject    handle to load_config_button (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB

```

```

% handles      structure with handles and user data (see
GUIDATA)

handles.experiment =
get(handles.experiment_choice, 'String');
handles.experiment =
handles.experiment{get(handles.experiment_choice, 'Value')};
display(handles.experiment);

set(handles.load_config_button, 'Enable', 'off');
set(handles.experiment_choice, 'Enable', 'off');
set(handles.reset, 'Enable', 'on');
set(handles.com_port, 'Enable', 'on');
set(handles.com_open, 'Enable', 'on');
set(handles.bar_graph, 'Visible', 'on');
set(handles.s_no, 'Enable', 'on');
set(handles.s_no_v, 'Enable', 'on');
set(handles.model, 'Enable', 'on');
set(handles.model_v, 'Enable', 'on');
set(handles.firmware, 'Enable', 'on');
set(handles.firmware_v, 'Enable', 'on');
set(handles.num_leds, 'Enable', 'on');
set(handles.num_leds_v, 'Enable', 'on');
set(handles.num_seq, 'Enable', 'on');
set(handles.num_seq_v, 'Enable', 'on');

if strcmp(handles.experiment, 'CRISP') == 1
    display('Hey You reached the CRISP colony
correctly');

handles.choice=1;

fid = fopen('SSSI_CRISP60.ini', 'rt');
if(fid == -1)
    error(['ERROR - Could not open GUI
initialization file ', 'SSSI_CRISP60.ini']);
end
try
    %Get baud rate for serial communications
    nextline = fgetl(fid);
    if(findstr(nextline, 'BAUDRATE') ~= 1)
        error('ERROR - Improper format in
initialization file');
    end
    loc = findstr(nextline, '=');

```



```

        handles.SSSI_baud =
str2num(nextline(loc+1:length(nextline)));

        %Get number of leds
        nextline = fgetl(fid);
        if(findstr(nextline,'CRISPLEDS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        loc = findstr(nextline,'=');
        handles.num_led =
str2num(nextline(loc+1:length(nextline)));

        %Get number of CRISP sequences
        nextline = fgetl(fid);
        if(findstr(nextline,'CRISPSEQS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        loc = findstr(nextline,'=');
        handles.num_seqs_cr =
str2num(nextline(loc+1:length(nextline)));
%         display(handles.num_seqs_cr);
%         handles.num_seqs_cr=60;

        %Get saturation value for detector system
        nextline = fgetl(fid);
        if(findstr(nextline,'SATURATION') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        loc = findstr(nextline,'=');
        handles.saturation_value =
str2num(nextline(loc+1:length(nextline)));

        %Get x-axis values corresponding to each
sequence
        nextline = fgetl(fid);
        if(findstr(nextline,'XVALS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        loc = findstr(nextline,'=');
        handles.xvals =
str2num(nextline(loc+1:length(nextline)));
%         display(handles.xvals);

```

```

        if(length(handles.xvals) ~=
handles.num_seqs_cr)
            error('ERROR - Length of XVALS does not
match number of CRISP sequences');
        end

        %Get CRISP sequences
        nextline = fgetl(fid);
        if(findstr(nextline,'SEQS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        handles.seqs_cr =
zeros(handles.num_seqs_cr,handles.num_led);
        for i=1:handles.num_seqs_cr
            handles.seqs_cr(i,:) = str2num(fgetl(fid));
        %#ok<ST2NM>
        end
        fclose(fid);

        axes(handles.bar_graph);

        handles.bar_handles = zeros(5,1);%no reason for
having 5 handles
        handles.bar_handles(1) =
bar(handles.xvals,ones(handles.num_seqs_cr,1));
        set(handles.bar_handles(1), 'FaceColor',[0 0
1]) %blue
        %         hold;
        %         display(handles.seqs_cr);
        display('Hey exiting the CRISP colony
correctly');

        catch
            fclose(fid);
            error(lasterr,'modal');% need to see what this
statement does
        end
        handles.hinv = pinv(handles.seqs_cr);

        elseif strcmp(handles.experiment,'HADAMARD') == 1
            display('Hey you reached the HADAMARD colony
correctly');

            handles.choice=2;

```

```

fid = fopen('SSSIgui31.ini','rt');
if(fid == -1)
    error(['ERROR - Could not open GUI
initialization file ', 'searchgui2.ini']);
end
try
    %Get baud rate for serial communications
    nextline = fgetl(fid);
    if(findstr(nextline, 'BAUDRATE') ~= 1)
        error('ERROR - Improper format in
initialization file');
    end
    loc = findstr(nextline, '=');
    handles.SSSI_baud =
str2num(nextline(loc+1:length(nextline)));

    %Get number of leds
    nextline = fgetl(fid);
    if(findstr(nextline, 'HADAMARDLEDS') ~= 1)
        error('ERROR - Improper format in
initialization file');
    end
    loc = findstr(nextline, '=');
    handles.num_led =
str2num(nextline(loc+1:length(nextline)));

    %Get number of hadamard sequences
    nextline = fgetl(fid);
    if(findstr(nextline, 'HADAMARDSEQS') ~= 1)
        error('ERROR - Improper format in
initialization file');
    end
    loc = findstr(nextline, '=');
    handles.num_seqs =
str2num(nextline(loc+1:length(nextline)));

    %Get saturation value for detector system
    nextline = fgetl(fid);
    if(findstr(nextline, 'SATURATION') ~= 1)
        error('ERROR - Improper format in
initialization file');
    end
    loc = findstr(nextline, '=');
    handles.saturation_value =
str2num(nextline(loc+1:length(nextline)));

    %Get x-axis values corresponding to each led

```

```

        nextline = fgetl(fid);
        if(findstr(nextline,'XVALS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        loc = findstr(nextline,'=');
        handles.xvals =
str2num(nextline(loc+1:length(nextline)));

        %Get hadamard sequences
        nextline = fgetl(fid);
        if(findstr(nextline,'SEQS') ~= 1)
            error('ERROR - Improper format in
initialization file');
        end
        handles.seqs =
zeros(handles.num_seqs,handles.num_led);
        for i=1:handles.num_seqs
            handles.seqs(i,:) = str2num(fgetl(fid));
        end

        fclose(fid);

    catch
        fclose(fid);
        error(lasterr,'modal');% need to see what this
statement does
    end

    % Find pseudo-inverse of hadamard transform
    handles.hinv = pinv(handles.seqs);
    display(handles.hinv);
    %     display(handles.xvals);

    axes(handles.bar_graph);

    %hold on;

    handles.bar_handles = zeros(5,1);
    handles.bar_handles(1) =
bar(handles.xvals(1:5),ones(5,1));
    set(handles.bar_handles(1), 'FaceColor',[0 0 1])
%blue
    hold on;

```

```

        handles.bar_handles(2) =
bar(handles.xvals(6:10),ones(5,1));
        set(handles.bar_handles(2), 'FaceColor', [0 1 0])
%green
        handles.bar_handles(3) =
bar(handles.xvals(11:15),ones(5,1));
        set(handles.bar_handles(3), 'FaceColor', [1 1 0])
%yellow
        handles.bar_handles(4) =
bar(handles.xvals(16:20),ones(5,1));
        set(handles.bar_handles(4), 'FaceColor', [1 0.5 0])
%orange
        handles.bar_handles(5) =
bar(handles.xvals(21:25),ones(5,1));
        set(handles.bar_handles(5), 'FaceColor', [1 0 0])
%red

        set(gca, 'xtick', [1,2,3,4,5]);

        hold off;

        display('Hey I am exiting the HADAMARD choice');
    else
        display('Hey something went wrong... find out
what!!!');
    end

    guidata(hObject,handles);

end

% --- Executes on selection change in com_port.
function com_port_Callback(hObject, eventdata, handles)
% hObject    handle to com_port (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

% Hints: contents = get(hObject,'String') returns com_port
contents as cell array
%         contents{get(hObject,'Value')} returns selected
item from com_port
end

```

```

% --- Executes during object creation, after setting all
properties.
function com_port_CreateFcn(hObject, eventdata, handles)
% hObject    handle to com_port (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    empty - handles not created until after all
CreateFcns called

% Hint: popupmenu controls usually have a white background
on Windows.
%         See ISPC and COMPUTER.
    if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
        set(hObject,'BackgroundColor','white');
    end

end

% --- Executes on button press in com_open.
function com_open_Callback(hObject, eventdata, handles)
% hObject    handle to com_open (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

% Get specified serial port from popup menu
com_port = get(handles.com_port,'String');
handles.SSSI_port =
com_port{get(handles.com_port,'Value')};
% Create serial object and open communications
handles.SSSI =
serial(handles.SSSI_port,'BaudRate',9600,'Timeout',10);
try
    fopen(handles.SSSI);
catch
    errordlg(['ERROR - Could not open communications on
',handles.SSSI_port, '. Matlab returned error:
',lasterr'],'modal');
    delete(handles.SSSI);
    return;
end

set(handles.SSSI,'Terminator',59);

handles.SSSI_open = 1;

```

```

set(handles.com_close, 'Enable', 'on');
set(handles.com_open, 'Enable', 'off');

% Test communications to see if SSSI is connected & ready
to start from the
% begining or needs to be reset

fwrite(handles.SSSI, 's');
comm = 0;
while(handles.SSSI.BytesAvailable < 25)
    pause(0.01);
    comm=comm+1;
    if comm == 500,
        display(handles.SSSI.BytesAvailable);
        %
        reset_gui(hObject, eventdata, handles)
        errorldg('Communication Error - Please verify that
SSSI is connected properly and turned on.It might also need
to be reset', 'Communication Error');
        return
    end
end

SSSI_reply = fscanf(handles.SSSI, '%s');
display(SSSI_reply);

%get the Serial number of SSSI device connected
handles.sno = str2num(SSSI_reply(2:5));
%display(handles.sno);
set(handles.s_no_v, 'String', handles.sno);

%get the Model Version of SSSI device connected
handles.mv = (SSSI_reply(6:7));
%display(handles.mv);
set(handles.model_v, 'String', handles.mv);

%get the Firmware version of SSSI device connected
handles.fv = (SSSI_reply(8:10));
%display(handles.fv);
set(handles.firmware_v, 'String', handles.fv);

%get the number of LEDs in the SSSI device connected
handles.snled = str2num(SSSI_reply(11:14));
%display(handles.snled);
set(handles.num_leds_v, 'String', handles.snled);

%get the number of Sequences in the SSSI device connected
handles.snseq = str2num(SSSI_reply(15:18));

```

```

%display(handles.snseq);
set(handles.num_seq_v, 'String', handles.snseq);

%get the Ready 'r' or 'R' signal the SSSI device connected
handles.SSSI_status = (SSSI_reply(19));
display(handles.SSSI_status);
%display(SSSI_reply(25));
%display('Now close the connection');

set(handles.single_scan, 'Enable', 'on');
set(handles.r_n_scans, 'Enable', 'on');
set(handles.file_name, 'Enable', 'on');
set(handles.save_scan, 'Enable', 'on');

guidata(hObject, handles);

end

% --- Executes on button press in com_close.
function com_close_Callback(hObject, eventdata, handles)
% hObject    handle to com_close (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

    fclose(handles.SSSI);
    set(handles.com_open, 'Enable', 'on');
    set(handles.com_close, 'Enable', 'off');

    % display('total count is');
    % display(handles.hd_scount);

    hfid = fopen('hadamard_scans.txt' , 'wt');
    if(hfid == -1)
        error(['ERROR - Could not create output file
', 'scan.txt']);
    end

    for i=1:(handles.hd_scount-1)
        fprintf(hfid, ' scan %d', i);
        % display('i wrote the scan number');
        for j=1:handles.num_leds
            fprintf(hfid, ' %d', handles.hd_scan(i,j));
        end
        % display('i wrote the scan values');

```



```

        fprintf(hfid, ' \n \n');
    end

    fclose(hfid);

    cfid = fopen('CRISP_scans.txt' , 'wt');
    if(cfid == -1)
        error(['ERROR - Could not create output file
', 'scan.txt']);
    end
    %   fprintf(cfid, ' CRISP scans will eventually be stored
here!!!');
    for i=1:(handles.cr_scount-1)
        fprintf(hfid, ' scan %d', i);
        %   display('i wrote the scan number');
        for j=1:60
            fprintf(hfid, ' %d', handles.cr_scan(i,j));
        end
        %   display('i wrote the scan values');
        fprintf(hfid, ' \n \n');
    end
    fclose(cfid);

    set(handles.single_scan, 'Enable', 'off');
    set(handles.r_n_scans, 'Enable', 'off');
    set(handles.file_name, 'Enable', 'off');
    set(handles.save_scan, 'Enable', 'off');

end

% --- Executes on button press in single_scan.
function single_scan_Callback(hObject, eventdata, handles)
% hObject    handle to single_scan (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)
    handles.totscans = 1;

    set(handles.single_scan, 'Enable', 'off');
    set(handles.r_n_scans, 'Enable', 'off');
    scan(hObject, eventdata, handles);
    set(handles.single_scan, 'Enable', 'on');
    set(handles.r_n_scans, 'Enable', 'on');

```

```

end

% --- Executes on button press in n_scans.
function n_scans_Callback(hObject, eventdata, handles)
% hObject    handle to n_scans (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

    %set(handles.single_scan,'Enable','off');
    set(handles.r_n_scans,'Enable','off');
    set(handles.num_scans,'Enable','off');
    set(handles.n_scans,'Enable','off');
    handles.totscans =
str2num(get(handles.num_scans,'String'));
%     handles.totscans =
str2num(sc_num{get(handles.num_scans,'Value')}});str2double(
get(hObject,'String'))

%     display(handles.totscans);

    if handles.totscans > 0

%         for i = 1 : handles.totscans
%             scan(hObject, eventdata, handles);
%         end
        scan(hObject, eventdata, handles);
    else
        display('Put in a valid number please');
    end

    set(handles.r_n_scans,'Enable','on');
    set(handles.num_scans,'Enable','on');
    set(handles.n_scans,'Enable','on');

end

function file_name_Callback(hObject, eventdata, handles)
% hObject    handle to file_name (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

```

```

% Hints: get(hObject,'String') returns contents of
file_name as text
%         str2double(get(hObject,'String')) returns contents
of file_name as a double
end

% --- Executes during object creation, after setting all
properties.
function file_name_CreateFcn(hObject, eventdata, handles)
% hObject    handle to file_name (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    empty - handles not created until after all
CreateFcns called

% Hint: edit controls usually have a white background on
Windows.
%         See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
end

% --- Executes on button press in save_scan.
function save_scan_Callback(hObject, eventdata, handles)
% hObject    handle to save_scan (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)
end

function num_scans_Callback(hObject, eventdata, handles)
% hObject    handle to num_scans (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

% Hints: get(hObject,'String') returns contents of
num_scans as text
%         str2double(get(hObject,'String')) returns contents
of num_scans as a double
end

```

```

% --- Executes during object creation, after setting all
properties.
function num_scans_CreateFcn(hObject, eventdata, handles)
% hObject    handle to num_scans (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    empty - handles not created until after all
CreateFcns called

% Hint: edit controls usually have a white background on
Windows.
%         See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
end

% --- Executes on button press in r_n_scans.
function r_n_scans_Callback(hObject, eventdata, handles)
% hObject    handle to r_n_scans (see GCBO)
% eventdata  reserved - to be defined in a future version
of MATLAB
% handles    structure with handles and user data (see
GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of
r_n_scans

    r_button = get(handles.r_n_scans,'Value');

    display(r_button);

    if r_button==1
        set(handles.single_scan,'Enable','off')
        set(handles.num_scans,'Enable','on');
        set(handles.n_scans,'Enable','on');
    else
        set(handles.single_scan,'Enable','on')
        set(handles.num_scans,'Enable','off');
        set(handles.n_scans,'Enable','off');
    end

end

% --- Executes on button press in reset.

```

```

function reset_Callback(hObject, eventdata, handles)
% hObject      handle to reset (see GCBO)
% eventdata    reserved - to be defined in a future version
of MATLAB
% handles      structure with handles and user data (see
GUIDATA)

% initial Flags for GUI

handles.SSSI_open=0;
handles.choice=0;

axes(handles.bar_graph);
bar(0,0);

set(handles.load_config_button, 'Enable', 'on');
set(handles.experiment_choice, 'Enable', 'on');
set(handles.reset, 'Enable', 'off');
set(handles.com_port, 'Enable', 'off');
set(handles.com_open, 'Enable', 'off');
set(handles.bar_graph, 'Visible', 'off');
set(handles.s_no, 'Enable', 'off');
set(handles.s_no_v, 'Enable', 'off');
set(handles.model, 'Enable', 'off');
set(handles.model_v, 'Enable', 'off');
set(handles.firmware, 'Enable', 'off');
set(handles.firmware_v, 'Enable', 'off');
set(handles.num_leds, 'Enable', 'off');
set(handles.num_leds_v, 'Enable', 'off');
set(handles.num_seq, 'Enable', 'off');
set(handles.num_seq_v, 'Enable', 'off');

reset_gui(hObject, eventdata, handles);

end

%-----
-----%
%-----USER functions-----
-----%
%-----
-----%

% -----
-----

```

```

% --- Reset GUI variables and controls
function reset_gui(hObject, eventdata, handles)
% -----
% -----
% Check if search is open and close
if(handles.SSSI_open)
    fclose(handles.SSSI);
    delete(handles.SSSI);
end
% if(handles.saving_data)
%     fclose(handles.savefile_id);
%     delete(handles.savefile_id);
% end
% handles.search_id = [];
% handles.savefile_id = [];
handles.SSSI_open = 0;
% handles.saving_data = 0;
% handles.scans_saved = 1;
% Set GUI control defaults
set(handles.com_close, 'Enable', 'off');
set(handles.single_scan, 'Enable', 'off');
set(handles.n_scans, 'Enable', 'off');
set(handles.file_name, 'Enable', 'off');
set(handles.save_scan, 'Enable', 'off');
set(handles.r_n_scans, 'Enable', 'off');
set(handles.num_scans, 'Enable', 'off');
set(handles.details_(sssi, 'ForegroundColor', [1 0 0]));

% set(handles.statustxt, 'String', 'Closed');
% set(handles.statustxt, 'ForegroundColor', [1 0 0]);
% set(handles.fstatustxt, 'String', 'Closed');
% set(handles.fstatustxt, 'ForegroundColor', [1 0 0]);
% set(handles.scansedit, 'String', '0');
% Update handles structure
guidata(hObject, handles);

end

% -----
% -----
% --- instructs the SSSI to scan the compound sample with
sequences based
% on experiment choice - CRISP or HADAMARD
function scan(hObject, eventdata, handles)
% -----
% -----

```

```

if strcmp(handles.experiment, 'HADAMARD') == 1

    try

        display('going to perform the HAD exp');
        numscans = handles.totscans;

        choice = '1';

        data = zeros(handles.num_seqs, numscans);
        vals = zeros(handles.num_seqs);
    %         display(numscans);

        for n = 1:numscans
            fwrite(handles.SSSI, choice);
            comm = 0;
            while(handles.SSSI.BytesAvailable <5)
                pause(0.01);
                comm=comm+1;
                if comm == 500,

display(handles.SSSI.BytesAvailable);
                %reset_gui(hObject, eventdata,
handles)
                errordlg('Communication Error -
Please verify that SSSI is connected properly and turned
on.It might also need to be reset', 'Communication Error');
                return
            end
        end
        display('Bytes available are:');
        display(handles.SSSI.BytesAvailable);

        for i=1:handles.num_seqs
            SSSI_val = fscanf(handles.SSSI, '%s');
    %         display(SSSI_val);
            if(i==1)
                temp =
str2num(SSSI_val(2:(length(SSSI_val)-1))); %#ok<ST2NM>
            else
                temp =
str2num(SSSI_val(1:(length(SSSI_val)-1))); %#ok<ST2NM>
            end
    %         display(temp);
            if(temp ~= i)

```

```

                                error('1','ERROR - Invalid sequence
order output from SSSI. '); %#ok<CTPCT>
                                end

                                SSSI_val = fscanf(handles.SSSI, '%s');
                                temp =
str2num(SSSI_val(1:(length(SSSI_val)-1))); %#ok<ST2NM>
                                %
                                display(temp);

                                data(i,n) = double(temp);
                                end
                                end
                                %
                                display(data);
display(data);

                                datatoworkspace =handles.hinv*data;
                                datatoworkspace = datatoworkspace';
                                assignin('base','data',datatoworkspace);

                                % Average
                                for i = 1:handles.num_seqs
                                    for n = 1:numscans
                                        vals(i) = vals(i) + data(i,n);
                                    end
                                    vals(i) = vals(i)/numscans;
                                end
                                %
                                display(vals);
                                if(handles.SSSI.BytesAvailable > 0)
                                    dummy =
fread(handles.SSSI,handles.SSSI.BytesAvailable);
                                    display('Dude there is still stuff in the
buffer!!!! wierd!!!!!!');
                                    display(dummy);
                                end
                                catch
                                    errordlg(lasterr,'modal');
                                    return;
                                end

                                if(~isempty(find(vals >=
handles.saturation_value))) %#ok<EFIND>
                                    errordlg('ERROR - Signal saturation
encountered','modal');
                                    return;
                                end

```



```

        % if (handles.saving_data)
        %     handles.scans_saved = handles.scans_saved +
1;
        % end

        vals = handles.hinv*vals;
        % display(vals);
        for i = 1:handles.num_leds
            if vals(i,1) < 0
                vals(i,1) = 0;
            end
            %handles.yvals(handles.scans_saved, i) =
vals(i,1);
        end

        axes(handles.bar_graph);

        for i=1:5
            set(handles.bar_handles(i), 'YData', vals((i-
1)*5+1):(i*5));
        end
        tempvals = vals(:,1);
        tempvals = tempvals';

        handles.hd_scan(handles.hd_scount,:)=tempvals(1,1:25);
        handles.hd_scount=handles.hd_scount+1;

        %set(handles.scansedit, 'String', num2str(handles.scans_saved
));
        guidata(hObject,handles);

    else
        display('Scanning');

        try

            display('going to perform the CRISP exp');
            numscans = handles.totscans;

            choice = '2';

            data = zeros(handles.num_seqs_cr, numscans);
            vals = zeros(handles.num_seqs_cr);
            %     display(numscans);

            for n = 1:numscans

```

```

        fwrite(handles.SSSI,choice);
        comm = 0;
        while(handles.SSSI.BytesAvailable <5)
            pause(0.01);
            comm=comm+1;
            if comm == 500,

display(handles.SSSI.BytesAvailable);
            %reset_gui(hObject, eventdata,
handles)
            errordlg('Communication Error -
Please verify that SSSI is connected properly and turned
on.It might also need to be reset', 'Communication Error');
            return
            end
        end
        %         display('Bytes available are:');
        %         display(handles.SSSI.BytesAvailable);

        for i=1:handles.num_seqs_cr
            SSSI_val = fscanf(handles.SSSI, '%s');
            %         display(SSSI_val);

            if(i==1)
                temp =
str2num(SSSI_val(2:(length(SSSI_val)-1)));
            else
                temp =
str2num(SSSI_val(1:(length(SSSI_val)-1)));
            end
            %         display(temp);
            if(temp ~= i)
                error('1', 'ERROR - Invalid sequence
order output from SSSI. '); %#ok<CTPCT>
            end

            SSSI_val = fscanf(handles.SSSI, '%s');
            temp =
str2num(SSSI_val(1:(length(SSSI_val)-1)));
            %         display(temp);

            data(i,n) = double(temp);

        end
    end
    display(handles.num_seqs_cr);
    % handles.hinv = pinv(handles.seqs_cr);

```

```

        display(data);

assignin('base','data',data');

    % Average
    for i = 1:handles.num_seqs_cr
        for n = 1:numscans
            vals(i) = vals(i) + data(i,n);
        end
        vals(i) = vals(i)/numscans;
    end
    display(vals);
    %
    if(handles.SSSI.BytesAvailable > 0)
        dummy =
fread(handles.SSSI,handles.SSSI.BytesAvailable);
        display('Dude there is still stuff in the
buffer!!!! wierd!!!!!!');
        display(dummy);
    end
    catch
        errordlg(lasterr,'modal');
        return;
    end

    if(~isempty(find(vals >=
handles.saturation_value))) %#ok<EFIND>
        errordlg('ERROR - Signal saturation
encountered','modal');
        return;
    end

    % if (handles.saving_data)
    %     handles.scans_saved = handles.scans_saved +
1;
    % end

    axes(handles.bar_graph);

set(handles.bar_handles(1),'YData',vals(1:handles.num_seqs_
cr));

    tempvals = vals(:,1);
    tempvals = tempvals';
    handles.cr_scan(handles.cr_scount,:)=tempvals(1,:);
    handles.cr_scount=handles.cr_scount+1;

```

```
%set(handles.scansedit,'String',num2str(handles.scans_saved
));
    guidata(hObject,handles);

    end
end
```

Appendix B: Solid State Spectral Imager User Manual

B.1 Instillation

Software Installation

The following software is needed to run the SSSI:

- MatLab version 7.0.1 or greater

The following software packages are needed for updating the SSSI user code:

- Kiel μ Vision3 V3.33
- Silicon Labs Integrated Development Environment. Version 3.10

Follow the recommended installation instructions provided with you software package.

Hardware Setup

The SSSI consists of three separate boards: emitter, controller, and detector. The three components combine to make a single unit. Figure B.1 illustrates the connection of the three boards. The SSSI requires a 12VDC battery for operation. Voltage regulators on each board are used to control the voltage for each component. The current prototype is powered on when connected to a 12VDC battery via power cables from each board. Care must be taken not to mix up the positive (red) and negative (green) power wires.

The SSSI is controlled via a serial connection made with a USB-serial adaptor to a computer. In order to update the C8051F processor (Silicon Laboratories Inc.) a JTAG/DEBUG cable is needed. The debug adaptor connects to the SSSI with the JTAG connector and to a computer using a USB connection.

B.2 Operation

SSSI GUI

- Connect the power cables to a 12VDC battery.
- Connect the SSSI to a computer running MatLab.
- Start MatLab
- Initialize the SSSI GUI
 - Find the MatLab directory containing the “SSSI_GUI.m” file and double click it.
 - The MatLab editor will display the GUI code. The GUI can be activated by pressing F5.
- When the SSSI GUI is active it will automatically be set for Hadamard scanning (see Figure B.2).
- Select the desired scanning mode using the drop down menu and load the configuration by pressing “load config”.
- Select the COM port that is connected to the SSSI using the drop-down menu in the GUI and press the “open” button. The SSSI should now be ready to scan.
- The SSSI can be operated in single scanning or multiple scanning modes. For single scanning press the single scan button. For multiple scanning select the n Scans radial button and enter the number of scans to acquire in the scans box. Press the n Scans button to begin acquiring multiple scan data.
- After each scan the data can be saved by entering a file name and clicking the “save scan” button. The data will be saved in a .txt file in the current directory.

- Data is automatically imported to the MatLab workspace after each scan. The variable “data” will contain the spectral data in the MatLab workspace.

Hadamard Scanning

Hadamard scanning is used to generate a sample spectrum and or multispectral image. The SSSI may be used to acquire spectroscopic images of a target via push broom scanning. However, this technique is not demonstrated in this manual.

- Initialize the SSSI GUI.
- Load the Hadamard configuration.
- The detector position is dependent on the scanning method. For transmission/absorption measurements the detector should be place 30cm in front of the collimator. Reflectance measurements can be made by placing the detector at a 45 degree angle from the target.
- Begin scanning.
- The SSSI GUI window will indicate the spectral response for each LED wavelength in the form of a bar graph (Figure B.3).

CRISP Scanning

CRISP scanning mode operation is the same for Hadamard scanning. Prior to initializing the SSSI GUI the CRISP sequence must be constructed. CRISP binary sequences can be generated with the “generateCRISP.m” file listed in Appendix A. The binary code

generated in MatLab must be converted to Hex before updating the processor code on the SSSI. This can be done using the “genhex.m” MatLab file.

CRISP Programming

- Start Kiel uVision software.
- Under the project tab press “open project”.
- Navigate to the directory containing the SSSI code and open HC602.uv2.
- The C-code (SSSI_v3.c) contains both the Hadamard and CRISP sequences. Scroll down to the CRISP SCAN Function routine. Currently there are 60 patterns which correspond to a CRISP scanning resolution of 10. The Hex code generated in MatLab must be copied to the C-code file. The MatLab generated Hex code is ordered according to its placement in the C-code file (see Figure B.4).
- Once the C-code is updated it must be saved.
- Under the project tab select “Rebuild All Target Files”. If the rebuild is successful the output window should indicate 0 Errors and 0 Warnings (Figure B.5).
- The code can now be uploaded to the SSSI.
- Connect the serial-USB and JTAG-USB cables to the SSSI and connect to a computer.
- Power the SSSI.
- Start Silicon Labs Integrated Development Environment. Version 3.10 for C8051xxxx processors.

- Connect to the SSSI by pressing the connect button and update the processor by pressing the “download code” key (see Figure B.6).
- After the code had been successfully updated, disconnect the SSSI by pressing the disconnect button (same as the connect button).
- Power the SSSI down and remove the JTAG cable.
- The final step is to update the SSSI.ini file in MatLab.
- Open MatLab and go to the directory that contains the SSSI GUI.
- Locate the SSSI_CRISP.ini file and open it in the MatLab editor.
- Paste the binary code into the .ini file under “SEQS”.
- Save the .ini file.
- Initialize the GUI and power the SSSI. CRISP data can now be acquired.

B.3 Troubleshooting

The SSSI GUI will not open:

- Check that the SSSI has power. Ensure there are no loose wires or connections
- An incorrect COM port is selected. The computer device manager in the system control panel will list all used ports. Ensure the proper COM port is selected
- Battery power is low. Check and replace battery if needed

The SSSI locked-up during a multiple scan:

- Close the SSSI GUI and reset

- MatLab may be restarted if a GUI reset does not work

The SSSI GUI does not respond:

- Close the GUI and restart
- Restart MatLab
- Close MatLab, power off the SSSI and disconnect it from the computer

Appendix B Figures

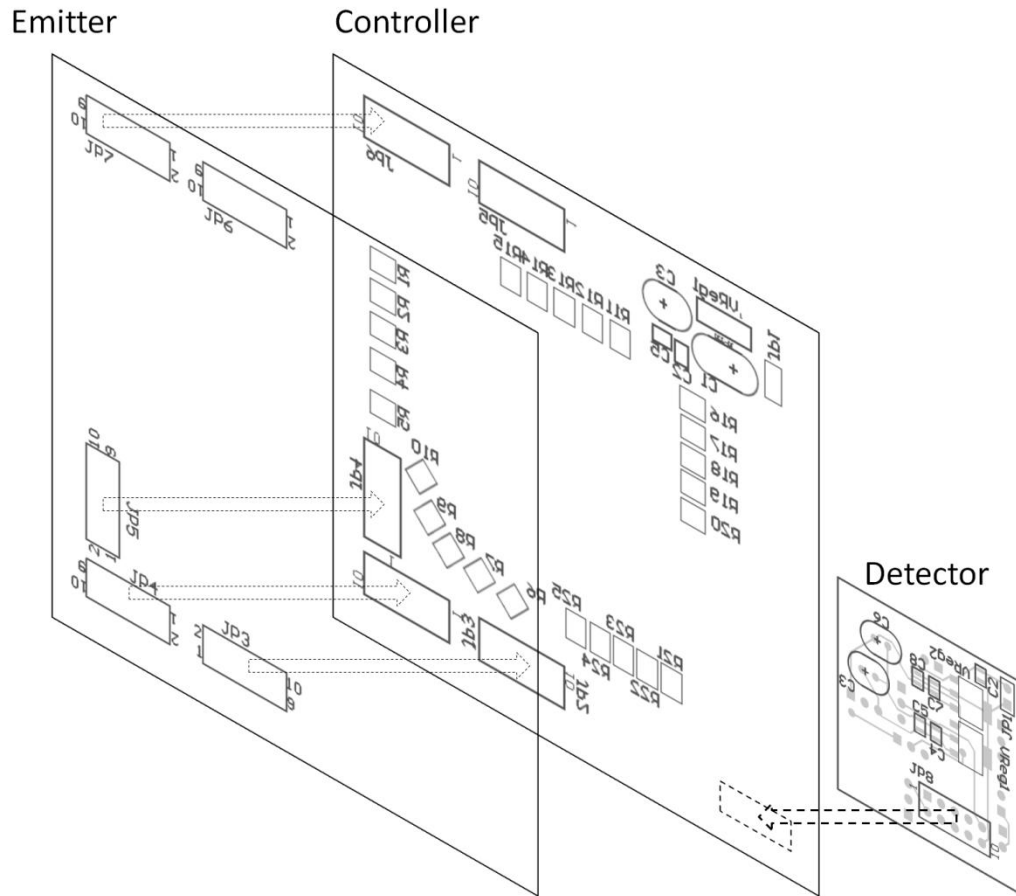


Figure B.1: Diagram of the connection sequence for the SSSI emitter, controller and detector boards.

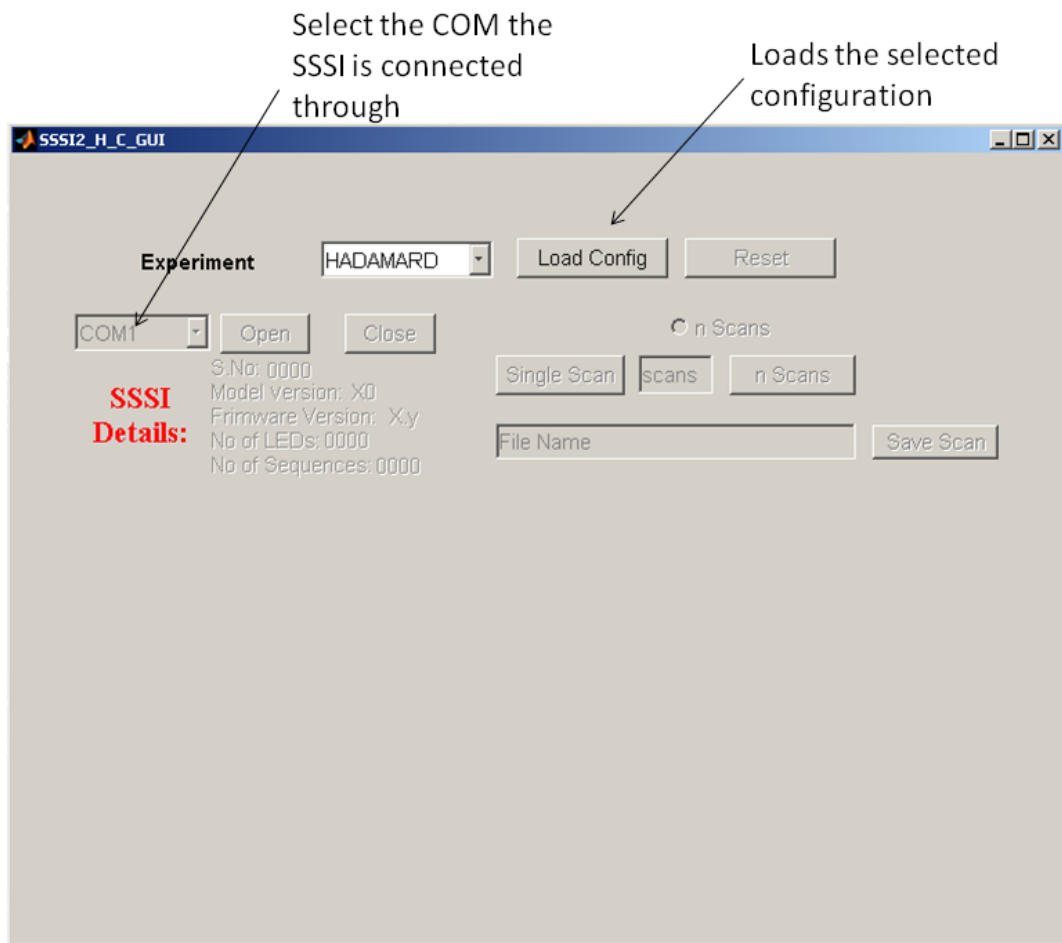


Figure B.2: Screen shot of the SSSI GUI.



Figure B.3: SSSI GUI in Hadamard scanning mode.

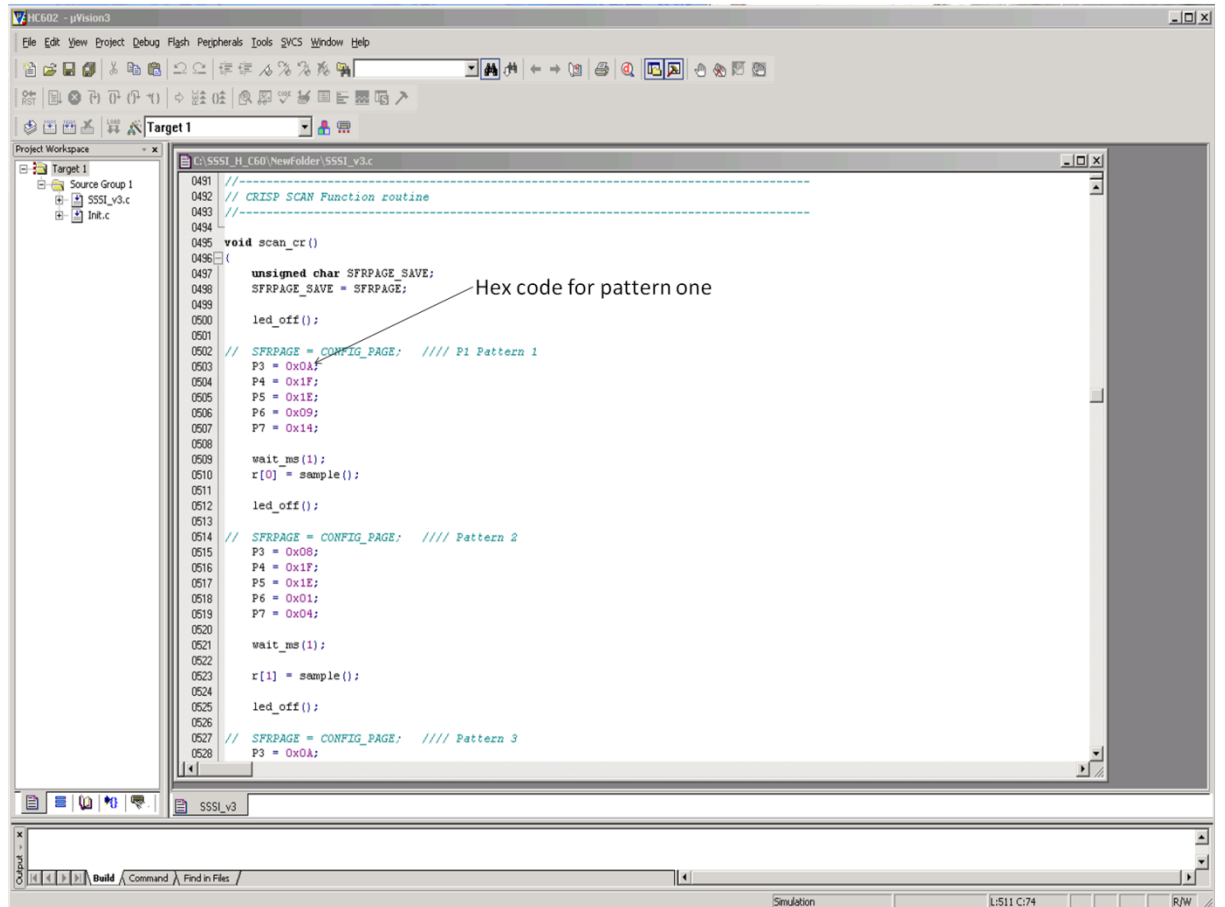


Figure B.4: Hex code must be copied into the C-code file for the SSSI. Each line of the Hex code corresponds to P3 to P7 for patterns 1 to 60.

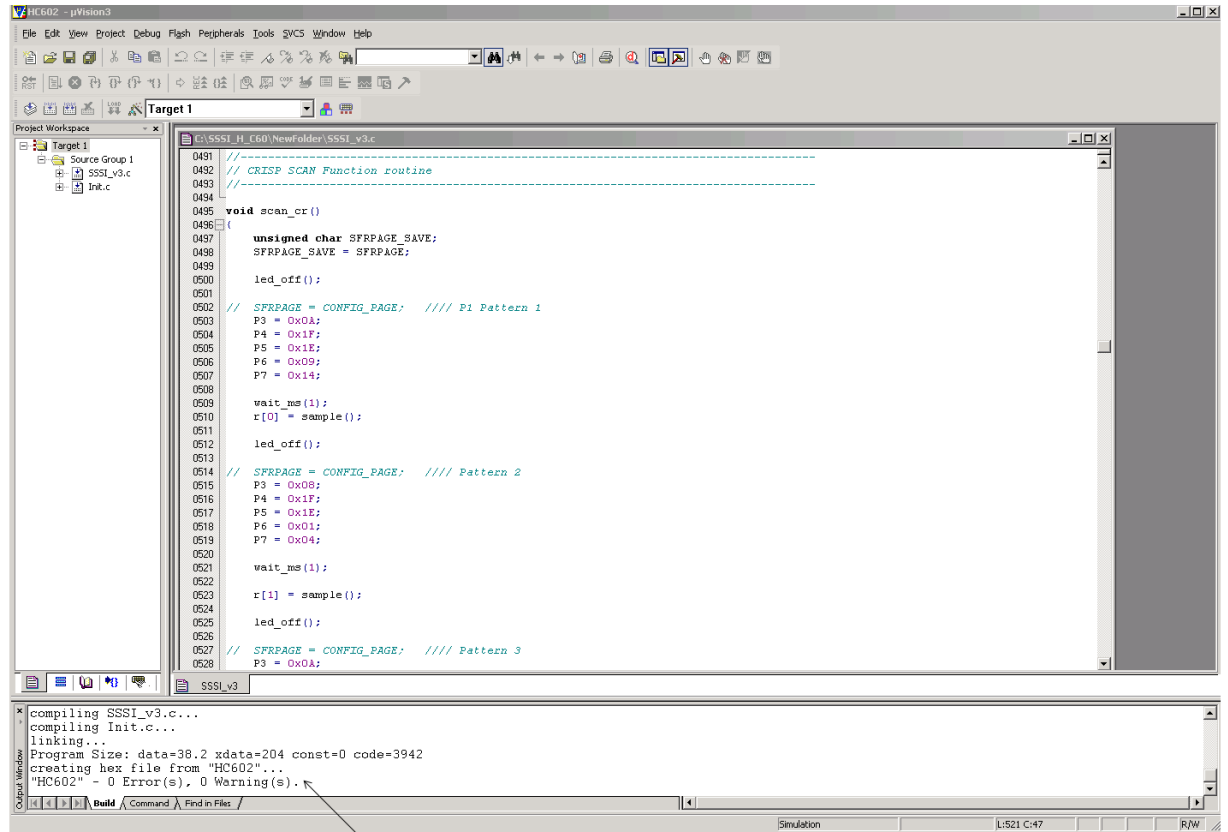


Figure B.5: The error and warning messages will indicate a successful rebuild of the target files.

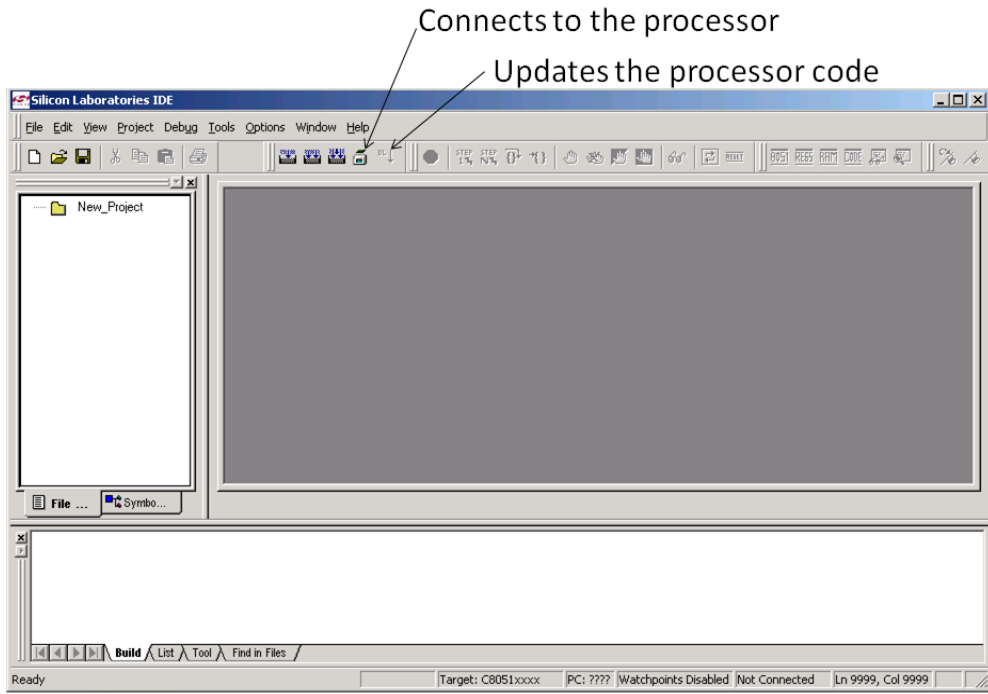


Figure B.6: The SSSI processor code is updated using Silicon Labs IDE software.

References

1. T. Hirschfeld, *Anal. Chem.* **1980**, 52, 297A-303A.
2. Sweedler, J.V. *Anal. Bioanal. Chem.* **2002**, 373, 321–322.
3. Ian D. Wilson, Udo A.Th. Brinkman, *Trends in Analytical Chemistry*, **2007**, 26(9), 847-854.
4. Su May Hsu, Hsiao-hua K. Burke, *Lincoln Laboratory Journal*, **2003**, 14(1), 145-159.
5. Lisa A. Cassis *et al.*, *Proc. SPIE-Int. Soc. Opt. Eng.* **2004**, 5329, 239-253.
6. DARPA Defense Sciences Office,
<http://www.darpa.mil/dso/thrusts/math/appliedcompmath/isp/index.htm>. Checked 9/26/2009
7. “Pattern recognition”, <http://dictionary.oed.com>. Checked 9/26/2009
8. Food and Drug Administration,
<http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm088828.htm>. Checked 10/2/2009
9. Hurricane Hunters Association, <http://www.hurricanehunters.com/>. Checked 9/26/2009.
10. D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York, **1992**.
11. S.J. Raudys and V. Pikelis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **1980**, 2, 243-251.
12. S.J. Raudys and A.K. Jain, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **1991**, 13(3), 252-264.

13. Anil K. Jain *et al.* IEE Transactions on Pattern Analysis and Machine Intelligence, **2000**, 22(1), 4-37.
14. Edward Jackson, A Users Guide to Principal Components, John Wiley & Sons, New York, **2003**
15. Miller, James N., Miller, Jane C., Statistics and Chemometrics for Analytical Chemistry 4th ed., Person Education Limited. **2000**, 217-221.
16. Trevor Hastie, *et al.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Ed. Springer, New York, **2009**.
17. David M. Haaland, Edward V. Thomas, *Anal. Chem.*, **1988**, 60(11), 1193–1202.
18. Tristrom Cooke, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2002**, 24(2), 268-273.
19. Food and Drug Administration,
<http://www.fda.gov/bbs/topics/NEWS/2006/NEW01507.html>. Checked 9/3/2007.
20. Sharaf, M. A.; Illman, D. L.; Kowalski, B. R., *Chemometrics*. Wiley: New York, **1986**.
21. Kowalski, B. R., *Chemometrics: Mathematics and Statistics in Chemistry*. Springer: New York, **2001**.
22. Brereton, R. G., *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, Inc.: Chichester, UK, **2003**.
23. Daniel P. Berrar, Werner Dubitzky, Martin Granzow, *A Practical Approach to Microarray Data Analysis*, Kluwer Academic Publishers, MA, **2003**.
24. Richard Hill, Elementary Linear Algebra with Applications, 3rd ED. Thomson, New York, **1996**.

25. Lodder, R. A.; Selby, M.; Hieftje, G. A., *Anal Chem* **1987**, 59, (15), 1921-1930.
26. Lodder, R. A.; Hieftje, G. A., *Applied Spectroscopy* **1988**, 42, (8), 1351-1365.
27. Lodder, R. A.; Hieftje, G. A., *Spectroscopy* **1988**, 42, (8), 1500-1512.
28. Zou, Y.; Xia, Y.; Jones, A. R.; Lodder, R. A., *Analytical Chemistry* **1993**, 65, (9), A434-A439.
29. Nair, Prakash; Lodder, Robert A., *Applied Spectroscopy* **1993**, 47(3), 287-291.
30. Efron, B., *Journal of the American Statistical Association* **2000**, 95(452), 1293-1296.
31. Efron, B., *Journal of the American Statistical Association* **1987**, 82(397), 171-185.
32. Diaconis, P.; Efron, B., *Scientific American* **1983**, 248(5), 116-121.
33. Efron, B.; Gong, G., *American Statistician* **1983**, 37(1), 36-48.
34. Food and Drug Administration, Guidance for Industry, Process Validation: General Principles and Practices, **2008**.
35. ICH, Pharmaceutical Quality System Q10,
<http://www.ich.org/LOB/media/MEDIA3917.pdf>. Checked 9/6/2009.
36. Food and Drug Administration,
<http://www.fda.gov/Safety/Recalls/ucm177321.htm>. Checked 9/6/2009.
37. Wild S, Roglic G, Green A, Sicree R, King H. *Diabetes Care*, **2004** 27, 1047–1053.
38. American Diabetes Association, *Diabetes Care*, **2008** 31, 596–615.

39. Agency for Health and Research Quality, Economic and Health Costs of Diabetes, <http://www.ahrq.gov/data/hcup/highlight1/high1.htm>. Checked 9/6/2009.
40. World Health Organization, <http://www.who.int/mediacentre/factsheets/fs312/en/>. Checked 9/6/2009.
41. UK Prospective Diabetes Study Group, *Lancet*, **1998**, 352, 837–853.
42. American Diabetes Association, *Diabetes Care*, **2003**, 26, S28 – S32.
43. Y. Lu, G. V. Levin and T. W. Donner, *Diabetes, Obesity and Metabolism*, **2008**, 10, 109–134.
44. Moore MC. *Curr Opin Invest Drugs*, **2006**, 7(10) 924–5.
45. Sara B. Police¹, J. Clay Harris, Robert A. Lodder and Lisa A. Cassis, *Obesity*, **2008**, doi:10.1038/oby.2008.508.
46. Spherix Inc., Spherix Incorporated Announces Positive Phase 2 Study Results. Press Release June 24, **2009**.
47. Baur, J. A. *et al. Nature*, **2006**, 444, 337–342.
48. Jang M. *et al. Science*, **1997**, 275, 218–220.
49. Jang DS. *et al. Biochem Pharmacol*, **1999**, 57, 705–712.
50. Rotondo S. *et al. Br J Pharmacol* , **1998**, 123, 1691–1699.
51. Bertelli A.A. *et al. Int J Tissue React*, **1995**, 17, 1–3.
52. Zang, M. *et al. Diabetes*, **2006**, 55, 2180–2191.
53. Hui-Chen Su, Li-Man Hung and Jan-Kan Chen. *Am. J. Physiol Endocrinol Metab*, **2006**, 290, 1339-1346.
54. Szkudelski, T. *Life Sciences*, **2008**, 82, 430–435.

55. L. Pari, M.A. Satheesh *Life Sciences*, **2006**, 79, 641–645.
56. M. Manickam *et al.* *J. Nat. Prod.* **1997**, 60, 609-610.
57. Buice Jr., R.G.; Pinkston, P.; Lodder, R.A. *App. Spec.* **1994**, 48(4), 517-524.
58. Medendorp, J.P.; Lodder, R.A. *AAPS PharmSciTech.* **2006**, 7(1), Article 25.
59. Hannel, T.; Link, D.; Lodder, R. *J. Pharm Innov.* **2008**, 3 (3), 152-160.
60. Martin, L.; Poret, J.; Danon, A.; Rosen, M. *Mat. Sci. Eng.* **1998**, A252, 27-35.
61. Bolotnikov, M.; Neruchev, Y. *J. Chem. Eng. Data.* **2003**, 48, 411-415.
62. Medendorp, J.P.; Buice Jr., R.G.; Lodder, R.A. *AAPS PharmSciTech.* **2006**, 7(3), Article 59.
63. Medendorp J., Lodder, R. *J. Chemometrics* **2005**; 19: 533–542
64. O. Soyemi, *et al.* *Anal. Chem.* **2001**, 73, 1069-1079.
65. Olusola O. *et al.* *Applied Spectroscopy.* **2002**, 56(4).
66. Bin Dai, *et al.* *Pharmaceutical Research*, **2007**, 24(8), 1441-1449.
67. Cassia, L.A.; Lodder, R. A., *Analytical Chemistry*, **1993**, 65, (9), 1247-1256.
68. Substance Abuse and Mental Health Services Administration,
<http://www.drugabusestatistics.samhsa.gov/NSDUH/2k6NSDUH/2k6results.cfm#Ch3>. Checked 2/29/2008.
69. Alcohol Epidemiologic Data System. Yi, H., Chen, C.M., and Williams, G.D. Surveillance Report #76: Trends in Alcohol-Related Fatal Traffic Crashes, United States, 1982–2004. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, Division of Epidemiology and Prevention Research (August 2006),
<http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/TrafficCrashes/crash01.htm>. Checked 2/29/2008.

70. NIH Publication No. 07-5010,
http://www.collegedrinkingprevention.gov/1College_Bulletin-508_361C4E.pdf,
Checked 2/29/2008.
71. NIH, <http://pubs.niaaa.nih.gov/publications/economic-2000/index.htm#updated>.
Checked 2/29/2008.
72. Substance Abuse and Mental Health Services Administration,
<http://www.oas.samhsa.gov/2k5/alcTX/alcTX.htm>. Checked 2/29/2008.
73. AW Jones, *et.al. Clinical Chemistry*, **1992**, 38, 743-747.
74. Rouen, David; Dolan, Kate. A Review of Drug Detection Testing and an
Examination of Urine, Hair, Saliva and Sweat. National Drug and Alcohol
Research Centre. ISBN 0-7334-1790-6, **2001**.
75. Katz, N. Fanciullo, GJ. *Clin. J. Pain.*, **2002**, 18(4), s76-82.
76. Hlastala, M. *J. of Appl. Physiol.*, **1998**, 84, 402-408.
77. R. J. Dempsey, D. G. Davis, R. G. Buice, Jr., R. A. Lodder; *Appl.*
Spectrosc., **1996**, 50(2), 18A-34A.
78. J. K. Drennen and R. A. Lodder. *J. Pharm. Sci.* **1990**, 79, 622–627.
79. A. S. El-Hagrasy, *et al.*, *J. Pharm. Sci.* **2001**, 90, 1298–1307.
80. A. Urbas, M. W. Manning, A. Daugherty, L. A. Cassis, and R. A. Lodder. *Anal.*
Chem., **2003**, 75, 3318–3323.
81. T. D. Ridder, S. P. Hendee, and C. D. Brown. *Appl. Spectrosc.*, **2005**, 59, 181–
189.
82. J. C. Soto, C. P. Meza, W. Caraballo, C. Conde, T. Li, K. R. Morris, and R. J.
Romanach. *Journal of Process Analytical Technology*, **2005**, 2(5), 8–14.

83. G Ingram, N Munro; *Br. J. Gen. Pract.*, **2005**, 55(516), 501–502.
84. Stephen F. Malin, Timothy L. Ruchti, Thomas B. Blank, Suresh N. Thennadil and Stephen L. Monfre, *Clinical Chemistry*, **1999**, 45, 1651-1658.
85. MR Robinson, RP Eaton, DM Haaland, GW Koepp, EV Thomas, BR Stallard and PL Robinson, *Clinical Chemistry*, **1992**, 38, 1618-1622.
86. Robbe C. Lyon, David S. Lester, E. Neil Lewis, Eunah Lee, Lawrence X. Yu, Everett H. Jefferson and Ajaz S. Hussain, *AAPS PharmSciTech*, **2002**, 3(3) article 17.
87. Burger J.; Geladi P., *J. Chemometrics*, **2005**, 19, 355–363.
88. Lisa A Cassis, Aaron Urbas, Robert A Lodder; *Analytical and Bioanalytical Chemistry*, Trends, **2004**, DOI: 10.1007/s00216-004-2979-1
89. H Hollien, G DeJong, C. A. Martin, R. Schwartz, K. Liljegren; *J. Acoust. Soc. Am.*, **2001**, 110(6), 3198-3206.
90. David A. Sousa, *How the Brain Learns to Read*, Corwin Press, **2005**.
91. J. C. Hill, G.Toffolon, *J. Stud. Alcohol*, **1990**, 51, 108–113.
92. Isaksson T, Kowalski B., *Appl. Spectrosc.* **1993**, 47(6), 702-709.
93. T F. Babor, J C. Higgins-Biddle, J B. Saunders, M G. Monteiro, *The Alcohol Use Disorders Identification Test 2nd ED*. World Health Organization, **2001**
94. Widmark, E.M.P. *Die theoretischen Grundlagen und die praktische verwendbarkeit der gerichtlich-medizinischen alkoholbestimmung*(The theoretical bases and the practical usability of the judicial-medical alcohol regulation). Berlin; Urban & Schwarzenburg; **1932**).
95. Thomas, E. V.; Haaland, A. C. *Anal. Chem.*, **1990**, 62, 1091-1099.

96. Beebe, K. R.; Kowalski, B. R. *Anal. Chem.*, **1987**, 59, 1007A-1017A.
97. Pratt, James. *Digital Image Processing*. 4th Ed. ISBN 978-0-471-76777-0, John Wiley & Sons, **2007**.
98. Adobe,
<http://kb.adobe.com/selfservice/viewContent.do?externalId=332261&sliceId=1>.
Checked 11/2/2008.
99. Hannel T, Lodder R.A., *ASRG Technical report*, **2009**.
100. Cassis, L. A.; Lodder, R. A., Near-IR imaging of atheromas in living arterial tissue. *Anal Chem* **1993**, 65, (9), 1247-56.
101. Dempsey, R. J.; Cassis, L. A.; Davis, D. G.; Lodder, R. A., Near-infrared imaging and spectroscopy in stroke research: lipoprotein distribution and disease. *Ann N Y Acad. Sci.* **1997**, 820, 149-69.
102. Moreno, P. R.; Lodder, R. A.; Purushothaman, K. R.; Charash, W. E.; O'Connor, W. N.; Muller, J. E., *Circulation* **2002**, 105, (8), 923-927.
103. F. G. Haibach, A. E. Greer, M. V. Schiza, R. J. Priore, O. O. Soysmi, and M. L. Myrick. *Applied optics*, **2003**, 42, 1833-1838.
104. F. G. Haibach and M. L. Myrick. *Applied optics*, **2004**, 43: 2130-2140.
105. M. L. Myrick, O. Soyemi, J. Karunamuni, D. Eastwood, H. Li, L. Zhang, A. E., Greer, and P. Gemperline. *Vibrational Spectroscopy*, **2002**, 28, 73-8.
106. M. L. Myrick, O. Soyemi, H. Li, L. Zhang, and D. Eastwood. *Fresenius' Journal of Analytical Chemistry*, **2001**, 369: 351-355.

107. M. L. Myrick, O. O. Soyemi, F. Haibach, L. Zhang, A. Greer, H. Li, R. Priore, M. V. Schiza, and J. R. Farr. *Proceedings of SPIE-The International Society for Optical Engineering*, **2002**, 4577, 148-157.
108. M. L. Myrick, O. O. Soyemi, M. V. Schiza, J. R. Farr, F. Haibach, A. Greer, H. Li, and R. Priore. *Proceedings of SPIE-The International Society for Optical Engineering*, **2002**, 4574, 208-215.
109. O. O. Soyemi, F. G. Haibach, P. J. Gemperline, and M. L. Myrick. *Applied Spectroscopy*, **2002**, 56, 477-487.
110. O. O. Soyemi, F. G. Haibach, P. J. Gemperline, and M. L. Myrick. *Applied Optics*, **2002**, 41, 1936-1941.
111. M.P. Nelson, J.F. Aust, J.A. Dobrowolski, P.G. Verly, M.L. Myrick, *Anal. Chem.*, **1998**, 70, 73-82.
112. L. A. Cassis, A. Urbas, and R. A. Lodder. *Anal. Bioanal. Chem.* **2005**, 382: 868–872.
113. NASA, <http://astrobiology.nasa.gov/about-astrobiology/>, 10/2/2009.
114. David J. Des Marais, *et. al.*, *Astrobiology*, **2008**, 8(4), 715-730.
115. Klein H.P., *Icarus*, **1978**, 34, 666-674.
116. Bell J.F., *J. Geophys. Res.*, **2000**, 105, 1721–1756.5.
117. Sizemore H.G., *et al.*, *Icarus*, **2009**, 199, 303-309.
118. A. Deanne Rogers, Joshua L. Bandfield, *Icarus*, **2009**, 203, 437-453.
119. C. P. McKay, *Origins Life Evol. Biosphere*, **1997**, 27, 263-289.
120. Victor R. Baker, *Nature* **2001**, 412, 228-236.

121. R.J. Dempsey, D.G. Davis, R.G. Buice, Jr., R.A. Lodder, *Appl. Spectrosc.*, **1996**, 50(2), 18A-34A.
122. Urbas A., *et al*, *Anal. Chem.* **2003**, 75, 3650-3655.
123. Carney J.M, *et al.*, *Anal. Chem.* **1993**, 65, 1305-1313.
124. Edwards HGM, Wynn-Williams DD, Villar SEJ, *J. Raman Spectroscopy*, **2004** 35(6), 470–474.
125. Villar SEJ, Edwards HGM, Cockell CS, *Analyst*, **2005**, 130(2), 156–162.
126. Villar SEJ, Edwards HGM , *Anal. Bioanal Chem.*, **2006**, 384, 100–113.
127. Organic Materials Review Institute, <http://www.omri.org/coppersulfate.pdf>, 9/17/2009
128. Stehouwer, R., Roth, G., *Field Crop News*, **2004**, 4(1).
129. Kamrin, M. A. Pesticide Profiles - Toxicity Environmental Impact, and Fate. CRC- Lewis Publishers, Boca Raton FL. **1997**, 421-578.
130. E. Hand, *Nature News*. **2008**, 10, 16.
131. M. H. Hecht, *et.al*. *Science*, **2009**, 325 (5936) 64-67.
132. Noriko Takamura, Fumie Kasai & Makoto M. Watanabe; *Journal of Applied Phycology*, **1989**, 1, 39-52.
133. J. D. Coates, L. A. Achenbach, *Nat. Rev. Microbiol.* **2004**, 2, 569-580.
134. Treado, P. J.; Morris, M. D. *Anal Chem* 1989; 61, 723A-734A.
135. Dieter, W., Lodder, R., Lumpp, J., IEEEAC paper #1439, **2004**, 3.
136. T. Hannel, R. Lodder, *Society of Amateur Radio Astronomers, Conference Proceedings*, **2008**, 95-104.

137. Green, R. O., Conel, J. E., and Chrien, T. G., Airborne Visible-Infrared Imaging Spectrometer (AVIRIS): Sensor System, In-flight Calibration and Reflection Calculation, Jet Propulsion Laboratory. ISSSR'92 Proceedings, **1992**, 1, pp198.
138. Clark, R. N., G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent. *J. Geophys. Res.*, **2003**, 108(E12), 5131, doi:10.1029/2002JE001847.
139. Andrew Lawler, *Science*, **2009**, 324. (5923), 34 – 35.
140. Roger C. Wiens, Sylvestre Maurice, CHEMCAM Laser-Induced Remote Sensing for Chemistry and Micro-Imaging, http://libs.lanl.gov/CHEMCAM_1_pg_web.pdf checked August 20, 2009.
141. Harald Furnes, Neil R. Banerjee, Karlis Muehlenbachs, Hubert Staudigel, Maarten de Wit, Early Life Recorded in Archean Pillow Lavas, *Science*, **2004**, 304 (5670), 578-581.
142. Gernot Arp, Andreas Reimer, Joachim Reitner. *Science* **2001**, 292, 1701-1704.
143. D.C. Catling, *Nature*, **2004**, 429, 707-708.
144. Olson JM. *Photosyn. Res.* **2006**, 88(2), 109–17.
145. Robert A. Ryerson, Andrew N. Rencz, Manual of Remote Sensing: Remote sensing for the earth sciences, John Wiley & Sons Inc. 1999.
146. McFadden L.A. et al., *Meteorites and Planetary Science*, 2001, 36, 1711-1726
147. R.A. Viscarra Rossel et al. *Geoderma*, **2006**, 131, 59–75.
148. Junfeng, J. I. et al. *Clays and Clay Minerals*, **2002**, 50(2), 208–216.
149. Lucey, P. G. et al. *J. Geophys Res*, **1986**, 91, D344-D354

150. Douglas, C. C.; et.al, In *ICCS, Part III, LNCS 3993*; V.N.Alexandrov et al. (Eds.); Springer-Verlag: Berlin Heidelberg, 2006; pp 393-400.
151. Sergios Theodoridis, Konstantinos Koutroumbas, *Pattern Recognition*, 3rd ED., Academic Press. 2006, 290
152. Hannel, T.S.; Sridhar, S.M.; Lodder, R.A. Complementary Integrated Sensing and Processing Solid State Spectroscopy (CRISP-SSSI) for Rapid Copper Sulfate Detection. In preparation for submission to *Applied Spectroscopy*.
153. USGS, <http://pubs.usgs.gov/of/2003/ofr-03-395/datatable.html>. Checked 9/1/2008.
154. John R. Rice, *Contact in Context*, **2001**. See also:
<http://cic.setileague.org/cic/v1i1/etcomeandgone.htm>.
155. Sabine C. Bauer, *Stargate SG-1: Trial by Fire: SG1-1*, Fandemonium Books; Mti edition 2007.
156. SETI, <http://www.setileague.org/general/drake.htm>. Checked 1/2/2008.
157. Allen Tough, *Contact in Context*, **2002**.
158. Stephen Webb, *If the Universe Is Teeming with Aliens... Where Is Everybody? Fifty Solutions to Fermi's Paradox and the Problem of Extraterrestrial Life*, Springer; 1st ed., 2002
159. NASA, http://heasarc.gsfc.nasa.gov/docs/cosmic/nearest_star.html. Checked 3/4/2008
160. General Atomics, *Nuclear Pulse Space Vehicle Study, Volume I -- Summary*, September 19, 1964.

161. Kirby J. Meyer, <http://www.engr.psu.edu/antimatter/documents.html>. Checked 3/4/2008.
162. Rutherglen; C, Burke; P, Nano Lett., 7, (11), 2007) (Appenzeller, J.; Frank, D. J. *Appl. Phys. Lett.* **2004**, 84(10), 1771-1773.
163. Rosenblatt, S.; Lin, H.; Sazonova, V.; Tiwari, S.; McEuen, P. *Appl. Phys. Lett.* **2005**, 87, 15311.
164. F. Garcia-Pichel and R. W. Castenholz, Occurrence of UV-Absorbing, Mycosporine-Like Compounds among Cyanobacterial Isolates and an Estimate of Their Screening Capacity, *Appl. Environ. Microbiol.*, **1993**, 59(1), 163-169.
165. Dieter, W., Lodder, R., Lumpp, J., IEEEAC paper #1439, Version 3, December 10, 2004
166. Buice R, Pinkston P, Lodder R. *Appl Spectrosc.* **1994**, 48, 517-524.
167. Medendorp J, Lodder RA. Applications of Integrated Sensing and Processing (ISP) in Acoustic and Optical Spectroscopy. Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy; February 27-March 4, 2005; Orlando, FL
168. Sandia Corporation, <http://www.sandia.gov/mstc/>. Checked 3/3/2008
169. Keith Wilson, Michael Enoch, IEEE Communications Magazine, August 2000
170. Kleine-Ostmann, T., Pierz, K., Hein, G., Dawson, P., and Koch, M. *Electron. Lett.* **2004**, 40, 124–126.
171. Q. Chen, M. Tani, Zhiping Jiang, and X.-C. Zhang. *J. Opt. Soc. Am. B.* **2001**, 18(6), 823-831.
172. Evan Reed. *Physical Review Letters*, January 11, **2006** (abstract) (<http://link.aps.org/abstract/PRL/v96/e013904>)

173. NASA, <http://nmp.jpl.nasa.gov/>, 2008

Vita

Thaddaeus Hannel

Born: June 29, 1975

Louisville, KY

EDUCATION

- Indiana University Southeast, Bachelor of Science in Chemistry, Minor in Mathematics, May 2005

PROFESSIONAL

- Teaching Assistant, University of Kentucky Department of Chemistry, 2008-2009
- Research Assistant, University of Kentucky Department of Chemistry, 2007-2008
- Teaching Assistant, University of Kentucky Department of Chemistry, 2006-2007

HONORS

- First place in the Ohio Valley Affiliates of Life Sciences (OVALS) poster competition 2009

PUBLICATIONS

Thaddaeus Hannel, Robert A. Lodder, "BEST-SCARS for Process Validation of Experimental Formulations of a Novel Diabetes Drug", **2009**. Submitted to *JPI*

Thaddaeus Hannel, Robert A. Lodder. "Bootstrap Error-Adjusted Single-Sample Technique Complementary Randomized Integrated Sensing and Processing (BEST-CRISP) for Classification of Common Martian Minerals and Identification of False Samples" **2009**. In preparation for submission to *Anal. Chem.*

Thaddaeus Hannel, David Link, Robert A. Lodder, "Surreptitious Remote Sensing of Blood Alcohol Content: Near-Infrared Spectroscopic (NIRS) Imaging and Laser Speech Detection", **2009**. Submitted to *Anal. Chem.*

Thaddaeus Hannel, Subash Marri Sridhar, Robert A. Lodder, "Complementary Integrated Sensing and Processing Solid State Spectroscopy (CRISP-SSSI) for Rapid Copper Sulfate Detection" **2009**. Submitted to *Applied Spectroscopy*.

Link, D., Hannel, T., Lodder, R.A. "ARS with BENDS to Quantify D-tagatose Concentrations in Resveratrol Tablets", **2009**. Submitted to *AAPS PharmaSciTech*

Link, D; Hannel, T; Lodder, R.A. “Bootstrap Enhanced N-dimensional Deformation of Space (BENDS) for Calibration of Nonlinear Responses in Acoustic Resonance Spectroscopy (ARS)”, **2009**, Submitted to *Algorithms*.

Craig C. Douglas, Gundolf Haase, Thad Hannel, David Link, and Robert A. Lodder, Deng Li. “A Prototype for Detecting Defective Pills during Manufacturing”, *International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, **2009**

Thaddaeus Hannel, Robert Lodder, “Size Does Matter: Solid State Spectral Imager (SSSI) on a Chip and the Search for Extraterrestrial Life”, *The Society of Amateur Radio Astronomers, 2008 Conference Proceedings*, 95-104, **2008**.

Thaddaeus Hannel, David Link, Robert Lodder, “Integrated Sensing and Processing - Acoustic Resonance Spectrometry (ISP-ARS) in Differentiating D-Tagatose and Other Toll Manufactured Drugs”, *Journal of Pharmaceutical Innovation*, 3 (3), 152-160, **2008**. DOI 10.1007/s12247-008-9038-y

C. C. Douglas, T. Hannel, D. J. Link, R. A. Lodder. Incorrect or Defective Pill Detection Using a Dynamic Data-Driven Application System Paradigm, *Workshop of Cyber Physical Systems*. **2008**.

Thaddaeus S. Hannel, Emmanuel O. Out, Mark P. Jensen, “Thermochemistry of the Extraction of Bismuth(III) with Bis(2-ethylhexyl) Phosphoric and 2-Ethylhexylphenylphosphonic Acids”, *Solvent Extraction and Ion Exchange*, 25(2): 241-256, **2007**. DOI: 10.1080/07366290601169493

Presentations

T. Hannel, R. Lodder. "Bootstrap Error-adjusted Single-sample Technique Integrated Sensing and Processing (BEST-ISP) for False Sample Identification" The Federation of Analytical Chemistry and Spectroscopy Societies: FACSS, 22 Oct. 2009, Louisville, KY.

T. Hannel, D. Link, R. Lodder. "Naturlose: A New Diabetes and Obesity Control Drug", Ohio Valley Affiliates for Life Sciences (OVALS) conference. 2-3 April 2009, Dayton, OH.

T. Hannel, D. Link, R. Lodder. “Integrated Sensing and Processing Acoustic Resonance Spectroscopy (ISP-ARS) versus Near-Infrared Spectroscopy (NIRS) for the Quantification of D-Tagatose in Resveratrol Tablets”, 60th Pittsburgh Conference on Analytical and Applied Spectroscopy (Pittcon). 8-13 March 2009, Chicago, IL.

T. Hannel, R. Lodder. “Solid State Spectral Imager (SSSI) for the Detection of Dissolved Solids in Diverse Aqueous Environments”. 60th Pittsburgh Conference on Analytical and Applied Spectroscopy (Pittcon). 8-13 March 2009, Chicago, IL.

T. Hannel, D. Link, R. Lodder. "Acoustic Resonance Spectroscopy for the Quantitative Determination of D-Tagatose in Resveratrol Tablets" AAPS Annual Meeting and Exposition. 19 November 2008, Atlanta, GA.

T. Hannel, D. Link, R. Lodder. "Surreptitious Remote Sensing of Blood Alcohol Content: Molecular Factor Computing (MFC) Near-Infrared Spectroscopic (NIRS) Imaging and Laser Speech Detection" The Federation of Analytical Chemistry and Spectroscopy Societies: FACSS, 30 Sept. 2008, Reno, NV.

Thaddaeus Hannel, Robert Lodder, "Size Does Matter: Solid State Spectral Imager (SSSI) on a Chip and the Search for Extraterrestrial Life". The Society of Amateur Radio Astronomers, 1 July 2008, Green Bank, WV.

T. Hannel, D. Link, R. Lodder. "Surreptitious Sensing of Blood Alcohol Content: Remote Near-Infrared Spectroscopic (NIRS) Imaging and Laser Speech Detection", Clinical and Translational Science Spring Conference (CCTS) 3 June 2008, Lexington, KY.

Thaddaeus Hannel, David Link and Robert A. Lodder, "Surreptitious Sensing of Blood Alcohol Content: Remote Near-Infrared Spectroscopic (NIRS) Imaging and Laser Speech Detection", 4th Kentucky Innovation and Enterprise Conference (KIEC), 17 April 2008, Lexington, KY.

Thaddaeus Hannel, David Link, Robert Lodder; "Integrated Sensing and Processing-Acoustic Resonance Spectroscopy (ISP-ARS) in Tablet Identification" Ohio Valley Affiliates for Life Sciences (OVALS). 14 April 2008, Louisville, KY.

Thaddaeus Hannel, David Link, and Robert A. Lodder, "Integrated Sensing and Processing-Acoustic Resonance Spectrometry (ISP-ARS) in Differentiating D-Tagatose and other Toll Manufactured Drugs. 34th Annual Naff Symposium on Chemistry and Molecular Biology, 4 April, 2008, Lexington, KY.

M. Kokorowski, C. Neish, J. Benson, S. Desiano, C. Gifford, T. Hannel, W. Huang, B. Johns, K. Lichtenberg, R. Macke, G. Minelli, A. Poppe, B. Schmidt, S. Taniguchi, D. Thompson, and T. Balint (2007), *VEIL: A New Frontiers Class Mission Concept for Exploring Venus*, American Geophysical Union (AGU) Fall Meeting, Poster P53A-0993, December 14, 2007.

C. Neish, J. Benson, S. Desiano, C. Gifford, T. Hannel, W. Huang, B. Johns, M. Kokorowski, K. Lichtenberg, R. Macke, G. Minelli, A. Poppe, B. Schmidt, S. Taniguchi, D. Thompson, T. Balint. *VEIL (Venus Exploration In-Situ Landers): A New Frontiers Class Mission Design Concept*. 39th Meeting of the AAS Division of Planetary Sciences, Orlando, FL., 7-12 October 2007

C. Neish, J. Benson, S. Desiano, C. Gifford, T. Hannel, W. Huang, B. Johns, M. Kokorowski, K. Lichtenberg, R. Macke, G. Minelli, A. Poppe, B. Schmidt, S. Taniguchi, D. Thompson, T. Balint. Venus In-Situ Mission , NASA Planetary Science Summer School, August 6 – 10, 2007, Pasadena CA.

Thaddaeus S. Hannel, Emmanuel O. Otu, Thermodynamics of the Extraction of Bismuth with Organophosphorus Acids, Presented at the 19th National Conference on Undergraduate Research, April 21-23, 2005, Lexington, Virginia

Thaddaeus S. Hannel, Emmanuel O. Otu. EFFECT OF TEMPERATURE ON THE EXTRACTION OF BISMUTH WITH ORGANOPHOSPHORUS ACIDS, Presented at the 10th Annual Indiana University Undergraduate Research Conference Friday, February 11, 2005, Indianapolis. IN.