

Biosensor-based Alternatives for Food Safety Analysis

submitted by:

Geoffrey C. Green, B.Eng, M.A.Sc.

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
(OCIECE)

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada K1S 5B6

January 2012

© Copyright 2012
Geoffrey C. Green



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-87756-2

Our file Notre référence

ISBN: 978-0-494-87756-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The objective of this thesis was to investigate the feasibility and potential of using specific biosensor-based methods – electronic nose (e-nose) and Raman spectroscopy – as alternatives to conventional methods for food safety analysis. The principal shortcoming associated with conventional methods is the length of time required for an unequivocal result (5-7 days). Measurement systems were designed by taking raw sensor responses from off-the-shelf biosensor instruments, extracting meaningful features from these (using appropriately selected signal processing algorithms), and performing pattern classification (and validation) to assess their effectiveness in identifying the contaminant.

For bacteria (*E. coli* and *Listeria*) hosted in nutrient broth, e-nose based systems using metal oxide sensors (MOS) and fingerprint mass spectrometry were used to discriminate between bacteria at different concentrations with discrimination evident at levels down to 10^6 cells/mL. The capacity of these systems to identify bacteria at lower concentrations was hindered by a time-varying drift inherent to the nutrient broth.

Using individual bacterial colonies, a MOS-based e-nose measurement system was able to discriminate between four bacteria types (two *E. coli* strains, *Listeria* and *Enterococcus*). Multiple consecutive e-nose responses were employed to achieve classification accuracies exceeding 90%. A surface-enhanced Raman spectroscopy based measurement system employing wavelet-based pre-processing and feature extraction stages was developed to discriminate between species of *Listeria*, yielding classification accuracies of over 90%.

Motivated by the potential to incorporate e-nose technology in a smart home environment, a MOS-based system was used to measure food spoilage. The e-nose sensor

responses demonstrated smooth trajectories in both original and dimension-reduced feature spaces and the degree of spoilage (as measured by sample age) were found to correlate with extracted features. In order to guide future research in this area, a framework for incorporating odour sensing into a smart home for identification of problems and activities of daily living is proposed.

Because the sample types used in this thesis are available relatively early in the inspection process, the performance of these methods justifies further work to assess their potential in a pre-screening capacity. The use of confidence measures is also explored to identify which samples might be classified with confidence with the proposed methods.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr. Adrian Chan and Dr. Rafik Goubran for the guidance, motivation, and support that they provided throughout the duration of this project. Dr. Chan in particular has continually made himself available to me for detailed technical discussions and advice on short notice, and for this I am very grateful. His careful attention to detail when reviewing our research results and publications has resulted in a much higher quality of work than would otherwise be possible.

I would like to express my appreciation to the Faculty of Graduate and Postdoctoral Affairs along with the Department of Systems and Computer Engineering (SCE) at Carleton University and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support.

I would also like to thank Dr. Min Lin, Hanhong Dan, and Steven Luo at the Canadian Food Inspection Agency for their assistance on this project. Their extensive knowledge of the food safety application was invaluable to me, and their assistance in preparing the bacteria samples used throughout the thesis is greatly appreciated. In addition, there was a lot of assistance provided by the engineers at AlphaMOS during the course of the project and I am grateful for their efforts.

I must also pass along my appreciation to colleagues at the Ottawa Hospital Research Institute (my place of employment since 2008) for their continued support while I completed the PhD program, to my many fellow colleagues and friends in the SCE

program for their companionship and helpful discussions, and to my family abroad for their love and support over the years (especially my mother and my late father).

Finally, I am most grateful to my wife Katherine (whom I have leaned on heavily during the last several years) for her unwavering support during the course of this program. Particularly towards the end of the program (as my “Dr. Daddy” time increased substantially), she has lovingly and patiently looked after Timothy, Douglas, and Lucie by putting them to bed, taking them to lessons, fixing their meals and helping with homework when I was not able to be around. Thank you.

Table of Contents

Abstract	iii
Acknowledgements	v
Table of Contents	vii
List of Tables	xi
List of Figures	xii
List of Abbreviations	xvi
Chapter One	1
1.1 Motivation	1
1.2 Problem Statement	2
1.3 Thesis Objectives	6
1.4 Thesis Contributions	7
1.4.1 Contributions: Discrimination between <i>E. coli</i> and <i>Listeria</i> Bacteria at Varying Concentrations in Nutrient Broth with Multiple E-nose Modalities.....	8
1.4.2 Contributions: Discrimination between Single Colonies of <i>E. coli</i> , <i>Listeria</i> and <i>Enterococcus</i> Bacteria with a Metal Oxide Sensor-based Electronic Nose.....	10
1.4.3 Contributions: Discrimination between <i>Listeria</i> Species with Surface-Enhanced Raman Spectroscopy	11
1.4.4 Contributions: Electronic Nose-based Tracking of Spoilage in Real Food Samples (with possible applications to Smart Home)	12
Chapter Two	15
2.1 Conventional Methods	15
2.1.1 Enrichment, Colony Growth, and Purification	16
2.1.2 Confirmation Tests.....	17
2.1.3 Advantages and Disadvantages.....	18
2.2 Enhancements to Conventional Methods.....	19
2.2.1 Immunoassays.....	20
2.2.2 Genetic Methods	21
2.3 Biosensor Methods.....	22
2.3.1 Biosensors based on Optical Properties	24
2.3.2 Biosensors based on Mass / Acoustic Wave Properties	25
2.3.3 Biosensors based on Electrochemical Properties.....	26
2.4 Electronic Nose (E-nose) Biosensors	27
2.4.1 Principle of Operation.....	27
2.4.2 Electronic Nose for Identification of Bacteria	30
2.5 Raman Spectroscopy	33
2.5.1 Principle of Operation.....	33
2.5.2 Raman Spectroscopy for Identification of Bacteria.....	35
2.6 Summary and Conclusions.....	37
Chapter Three	38
3.1 Background	38
3.2 Methods.....	40

3.2.1	Bacteria Samples.....	40
3.2.2	Electronic Nose Instruments Used.....	40
3.2.3	Pre-processing and Feature Extraction	47
3.2.4	Dimensionality Reduction	48
3.2.5	Classification and Performance Characterization.....	51
3.3	Experiment #1 – Exploratory Analysis of Dense Cultures Using Three E-nose Modalities	52
3.3.1	Experiment #1: Bacteria Sample Preparation and Sampling Protocol	52
3.3.2	Experiment #1: Electronic Nose Processing.....	53
3.3.3	Experiment #1: Results.....	53
3.3.4	Experiment #1: Conclusions and Discussion.....	55
3.4	Experiment #2 – Discrimination of Bacteria with Varying Concentrations (Dilutions Done Prior to Sampling).....	55
3.4.1	Experiment #2: Bacteria Sample Preparation and Sampling Protocol	55
3.4.2	Experiment #2: Electronic Nose Processing.....	56
3.4.3	Experiment #2: Results.....	56
3.4.4	Experiment #2: Conclusions and Discussion.....	59
3.5	Experiment #3 – Discrimination of Bacteria with Varying Concentrations (Dilutions at t=0).....	62
3.5.1	Experiment #3: Bacteria Sample Preparation and Sampling Protocol	62
3.5.2	Experiment #3: Electronic Nose Processing.....	63
3.5.3	Experiment #3: Results.....	63
3.5.4	Experiment #3: Conclusions and Discussion.....	65
3.6	Overall Summary and Conclusions.....	65
Chapter Four	68
4.1	Background	68
4.2	Methods.....	71
4.2.1	Electronic Nose Analysis.....	71
4.2.2	Data Processing - Baseline manipulation	72
4.2.3	Data Processing - Feature Extraction and Dimensionality Reduction.....	72
4.2.4	Classification and Performance Characterization.....	74
4.2.5	Confidence in Classification Results	74
4.3	Experiment #1 – Initial Feasibility Study	78
4.3.1	Experiment #1: Bacteria Sample Preparation	78
4.3.2	Experiment #1: Sampling Protocol.....	79
4.3.3	Experiment #1: Results.....	80
4.3.4	Experiment #1: Conclusions and Discussion.....	88
4.4	Experiment #2 – Validation Study.....	89
4.4.1	Experiment #2: Bacteria Sample Preparation	89
4.4.2	Experiment #2: Sampling Protocol.....	90
4.4.3	Experiment #2: Results.....	92
4.4.4	Experiment #2: Conclusions and Discussion.....	102
4.5	Overall Summary and Conclusions.....	104

Chapter Five	107
5.1 Background	107
5.2 Methods.....	110
5.2.1 Sample Preparation	110
5.2.2 Raman Spectra Measurement	111
5.2.3 Raman Spectral Pre-processing	112
5.2.4 Feature Vector Extraction and Dimensionality Reduction	115
5.2.5 Classification and Performance Characterization.....	118
5.2.6 Feature Ranking.....	119
5.2.7 Confidence Measures.....	120
5.3 Results	120
5.4 Discussion and Conclusion	126
 Chapter Six	 128
6.1 Background	129
6.2 Methods.....	133
6.2.1 Electronic Nose Analysis.....	133
6.2.2 Data Processing – Artificial Neural Networks.....	133
6.2.3 Data Processing – Fuzzy Clustering	134
6.2.4 Data Processing – Dimensionality Reduction.....	135
6.3 Experiment #1 – Feasibility Study with Multiple Foods, Irregular Sampling. 135	
6.3.1 Experiment #1: Food Samples, Sample Preparation and Sampling Protocol	135
6.3.2 Experiment #1: Additional Details	137
6.3.3 Experiment #1: Results.....	137
6.3.4 Experiment #1: Conclusions and Discussion.....	141
6.4 Experiment #2 – Validation Study with Single Food and Regular Sampling.. 142	
6.4.1 Experiment #2: Food Samples, Sample Preparation and Sampling Protocol	142
6.4.2 Experiment #2: Additional Details	142
6.4.3 Experiment #2: Results.....	143
6.4.4 Experiment #2: Conclusions and Discussion.....	144
6.5 Experiment #3 – Validation Study with Regular Sampling from Same Vial and Varying Storage Temperatures	145
6.5.1 Experiment #3: Food Samples, Sample Preparation and Sampling Protocol	145
6.5.2 Experiment #3: Additional Details	146
6.5.3 Experiment #3: Classification.....	147
6.5.4 Experiment #3: Results.....	148
6.5.5 Experiment #3: Conclusions and Discussion.....	152
6.6 Integration of Odour Monitoring into a Smart Home: Proposed Framework.. 154	
6.6.1 Spatially Distributed E-nose Array.....	157
6.6.2 Wireless communication capability.....	157
6.6.3 Expert Systems.....	158
6.6.4 Multi-Expert Fusion Block	158
6.6.5 Alerting System	160

6.6.6	Practical Issues to Overcome	160
6.7	Overall Summary and Conclusions.....	161
Chapter Seven	164
7.1	Summary of Research	164
7.2	Summary of Research Results: Chapter 3.....	165
7.3	Summary of Research Results: Chapter 4.....	166
7.4	Summary of Research Results: Chapter 5.....	167
7.5	Summary of Research Results: Chapter 6.....	168
7.6	Conclusions	169
7.7	Suggestions for Future Research.....	171
References	177

List of Tables

Table 4.1: MOS e-nose parameters used in single colony bacteria discrimination experiments.....	72
Table 4.2: Experiment #1: Summary of sample numbers and category labels.....	79
Table 4.3: Experiment #1: Overall and pre-category classification accuracy, for various input feature combinations.....	81
Table 4.4: Experiment #1: Confusion matrix for three-category classification...	82
Table 4.5: Sampling protocol for Experiment #2.....	91
Table 4.6: Experiment #2: Confusion matrix for the five-category classification problem using LOO.....	93
Table 4.7: Experiment #2: Confusion matrix for the five-category classification problem using modified LOO.....	94
Table 4.8: Experiment #2: Confusion matrices for the five category classification (presented for each sniff).....	97
Table 4.9: Classification accuracies using composite e-nose responses using various subsets of sniffs.....	100
Table 4.10: Experiment #2: Confusion matrices for five category classification	101
Table 5.1: Summary of sample numbers and category labels used for experiment in this chapter.....	111
Table 5.2: Classification results and confusion matrix for <i>Listeria</i> identification	120
Table 5.3: Classification results for identification of <i>L. monocytogenes</i>	121
Table 5.4: Feature Ranking.....	121
Table 6.1: Experiment #1: Mean, SD, and median of correlation coefficients...	140
Table 6.2: Experiment #3: Summary of the number of samples used and smellprints generated.....	146
Table 6.3: Experiment #3: Classification accuracies for food and storage temperature combinations.....	152

List of Figures

Figure 1.1: Processing stages involved in conventional methods of foodborne pathogen detection.....	4
Figure 1.2: Simplified view of conventional methods, annotated with the thesis contributions and chapters in which each contribution is described	8
Figure 2.1: General representation of an example of a dichotomous tree used in bacteria identification.....	18
Figure 2.2: Main functional components in an e-nose measurement system.....	29
Figure 2.3: Illustration of energy levels associated with Raman spectroscopy.....	33
Figure 2.4: Main functional components in a Raman spectroscopy measurement system.....	35
Figure 3.1: A set of sensor response time curves for a MOS-based electronic nose.....	42
Figure 3.2: FMS response curve for an arbitrary <i>E. coli</i> sample.....	43
Figure 3.3: A representative set of sensor response time curves for a CPS-based electronic nose.....	45
Figure 3.4: Photograph of e-nose instruments.....	46
Figure 3.5: Illustration of the feature extraction process.....	48
Figure 3.6: Experiment #1: PCA plots.....	54
Figure 3.7: Experiment #2: Clustering results, MOS.....	57
Figure 3.8: Experiment #2: Clustering results, FMS.....	58
Figure 3.9: Experiment #2: Clustering results, pooled samples over all runs.....	61
Figure 3.10: Experiment #2: Clustering results, LB samples only.....	62
Figure 3.11: Experiment #3: Clustering results, all runs.....	64
Figure 3.12: Experiment #3: Clustering results, Run #3 only.....	64
Figure 3.13: Simplified view of conventional methods (Chapter 3 highlighted)...	67

Figure 4.1: Illustration of the manner in which multiple colonies can grow on a plate during food testing.....	70
Figure 4.2: Plot illustrating feature extraction (2 features).....	74
Figure 4.3: Demonstration of the calculation of the silhouette coefficient.....	77
Figure 4.4: Photograph of bacteria colonies grown on agar substrate.....	78
Figure 4.5: Experiment #1: Clustering results.....	81
Figure 4.6: Experiment #1: Illustration of confidence measures for classifier results.....	83
Figure 4.7: Experiment #1: Illustration of confidence measures for classifier results, specific samples.....	85
Figure 4.8: Experiment #1: Effect of varying confidence threshold on classification accuracy and rejection rate.....	86
Figure 4.9: Experiment #1: PCA results with samples further categorized by the run number.....	87
Figure 4.10: Experiment #1: ULDA results with samples further categorized by the run number.....	88
Figure 4.11: Experiment #2: Clustering results.....	92
Figure 4.12: Experiment #2: Effect of varying confidence threshold on classification accuracy and rejection rate.....	95
Figure 4.13: Experiment #2: Overall and per-category classification accuracy as a function of sniff number.....	96
Figure 4.14: Experiment #2: Features extracted from consecutive sniffs demonstrating “first sniff” effect.....	98
Figure 4.15: Simplified view of conventional methods (Chapter 4 highlighted)...	106
Figure 5.1: Scanning electron microscope image of SERS bacteria sample.....	109
Figure 5.2: Typical SERS spectra for <i>Salmonella</i> , <i>Escherichia</i> and six species of <i>Listeria</i>	110

Figure 5.3: Block diagram of overall Raman measurement system to identify <i>Listeria</i> species.....	112
Figure 5.4: Wavelet transform coefficients at different levels of decomposition...	114
Figure 5.5: Feature Extraction from SERS spectra.....	117
Figure 5.6: Effect of varying confidence threshold on classification accuracy and rejection rate.....	122
Figure 5.7: Effect of DR on classification accuracy (with bootstrap resampling)...	123
Figure 5.8: Clustering in DR space (six-category case).....	124
Figure 5.9: Effect of varying T_A	125
Figure 5.10: Simplified view of conventional methods (Chapter 5 highlighted)...	127
Figure 6.1: Photographs of food samples used in Experiment #1.....	136
Figure 6.2: Experiment #1: PCA clustering results for Day 1 and 7 food samples	138
Figure 6.3: Experiment #1: PCA clustering results for food samples measured throughout the entire experiment.....	139
Figure 6.4: Experiment #1: Results of regression analysis.....	139
Figure 6.5: Experiment #1: DR Representation of all food samples – all days, all categories.....	141
Figure 6.6: Experiment #2: PCA results for milk sampled for one week at 4h intervals.....	143
Figure 6.7: Experiment #2 : Results of regression analysis.....	144
Figure 6.8: Experiment #3: Time progression of feature vector components for yogurt.....	150
Figure 6.9: Experiment #3: Time progression of feature vector components for milk.....	150
Figure 6.10: Experiment #3: PCA representation of milk and yogurt e-nose responses.....	151
Figure 6.11: Experiment #3: FCM results.....	151

Figure 6.12: Key elements of a smart home environment system that includes odour monitoring..... 156

Figure 6.13: Simplified view of conventional methods (Chapter 6 highlighted)... 163

List of Abbreviations

ADL	Activities of Daily Living
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARC	Accuracy Rejection Curve
BHI	Brain Heart Infusion
CFIA	Canadian Food Inspection Agency
CFU	Colony Forming Units
CPS	Conducting Polymer Sensor
DNA	Deoxyribonucleic Acid
DR	Dimensionality Reduction
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
ELISA	Enzyme Linked Immunosorbent Assay
FCM	Fuzzy C-means Clustering
FIR	Finite Impulse Response
FMS	Fingerprint Mass Spectrometry
FTIR	Fourier-Transform Infrared Spectroscopy
GC	Gas Chromatography
HACCP	Hazard Analysis and Critical Control Point
HCA	Hierarchical Cluster Analysis
IEEE	Institute of Electrical and Electronics Engineers
LB	Luria-Bertani

LDA	Linear Discriminant Analysis
LOO	Leave One Out
MLP	Multilayer Perceptron
MOS	Metal Oxide Sensor
MS	Mass Spectrometry
NIR	Near Infrared
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PTFE	Polytetrafluoroethylene
QCM	Quartz Crystal Microbalance
SAW	Surface Acoustic Wave
SERS	Surface Enhanced Raman Spectroscopy
SOM	Self Organizing Maps
SPR	Surface Plasmon Resonance
SVM	Support Vector Machine
SWT	Stationary Wavelet Transform
TAFETA	Technology Assisted Friendly Environment for the Third Age
ULDA	Uncorrelated Linear Discriminant Analysis
UVRR	Ultraviolet Resonance Raman Spectroscopy
VBNC	Viable But Nonculturable
VOC	Volatile Organic Compounds
WHO	World Health Organization

Chapter One

Introduction

1.1 Motivation

The World Health Organization (WHO) has identified foodborne illnesses – defined as diseases that enter the body through ingestion of food – as one of the most serious public health threats facing the world today [1]. Foodborne illness is caused by a variety of pathogens such as viruses, bacteria, and parasites. Illness severity can range from the very mild (e.g., gastroenteritis) to the life threatening (e.g., hemolytic uremic syndrome, cholera, and listeriosis) [2].

While the overall safety of food has been considerably improved in recent decades, the incidence and severity of foodborne illness in humans remains unacceptably high worldwide. Consider the following statistics:

- Globally, it has been estimated by the WHO that over 1.8 million people die annually of diseases directly attributable to food and water contamination [1],
- In industrialized countries, around 30% of the population annually are affected by episode(s) of foodborne illness [1],
- In the United States, roughly 325 000 hospitalizations and 5 000 deaths occur annually due to foodborne illness [1],
- Though difficult to measure, the monetary cost associated with foodborne illness in the US has been estimated between tens of billions to over \$1 trillion annually [3][4], and

- The potential for outbreaks of illness to spread is increased by the large extent to which (possibly contaminated) foods are traded in the increasingly globalized economy [5].

There are population segments at increased risk of contracting foodborne illness. For instance, in the developing world, the problem is exacerbated by unsanitary food handling conditions, malnutrition, and inadequate or non-existent access to food inspection technology. In industrialized countries, older adults are more susceptible to foodborne illness. This has been attributed to decrease in cellular immunity, age-related changes in gastric activity and impaired cognitive abilities [6][7].

1.2 Problem Statement

The development of excellent programs and standards related to food manufacturing and handling (notably, ISO 22000:2005 and hazard analysis and critical control point (HACCP) systems [8][9]), have reduced the extent to which foodborne pathogens are found in the food supply. Despite this, it has proved impossible to eliminate them entirely. Consequently, methods for pathogen detection and identification continue to play a major role in food safety.

Conventional methods that are employed to recognize the presence and nature of foodborne pathogens rely on the techniques of *microbiological* and *biochemical identification*. Principally, there are three categories of methods that have been used by food inspection authorities [10]:

1. Culturing of pathogenic species, followed by colony counting,

2. Immunological-based (in which the organism's antibody-antigen interactions are used), and
3. Polymerase chain reaction (in which the organism's genetic information is analyzed).

These conventional approaches have been used for years – they have been largely successful, providing a battery of sensitive tests that can provide qualitative and quantitative information as to the presence, nature and extent of food contamination. However, their use is sub-optimal in several respects [10][11][12]:

1. There is generally a very *long time required* (ranging from a couple of days to over a week) before definitive results are available. This is due to the requirement for multiple culturing phases, or *enrichments*, which increase the number of organisms from a very low count to a dense culture (thereby increasing the pathogen's "signal-to-noise ratio" for subsequent analysis) ,
2. There is a *relatively high complexity* associated with these methods, which necessitates highly skilled laboratory technicians and significant labour and reagent expenses, and
3. The methods require clean laboratory facilities and thus, are not generally portable and unsuitable for *in situ* monitoring.

A simplified illustration of the steps required to identify the origin of a pathogenic species using conventional culturing methods is shown in Figure 1.1. Evident are multiple culturing phases, each requiring a significant incubation time (24-48 hours) to allow sufficient colony growth and purity for the identification tests, which occur at the

end of the process. A more detailed description of this process is provided in the next chapter (Section 2.1).

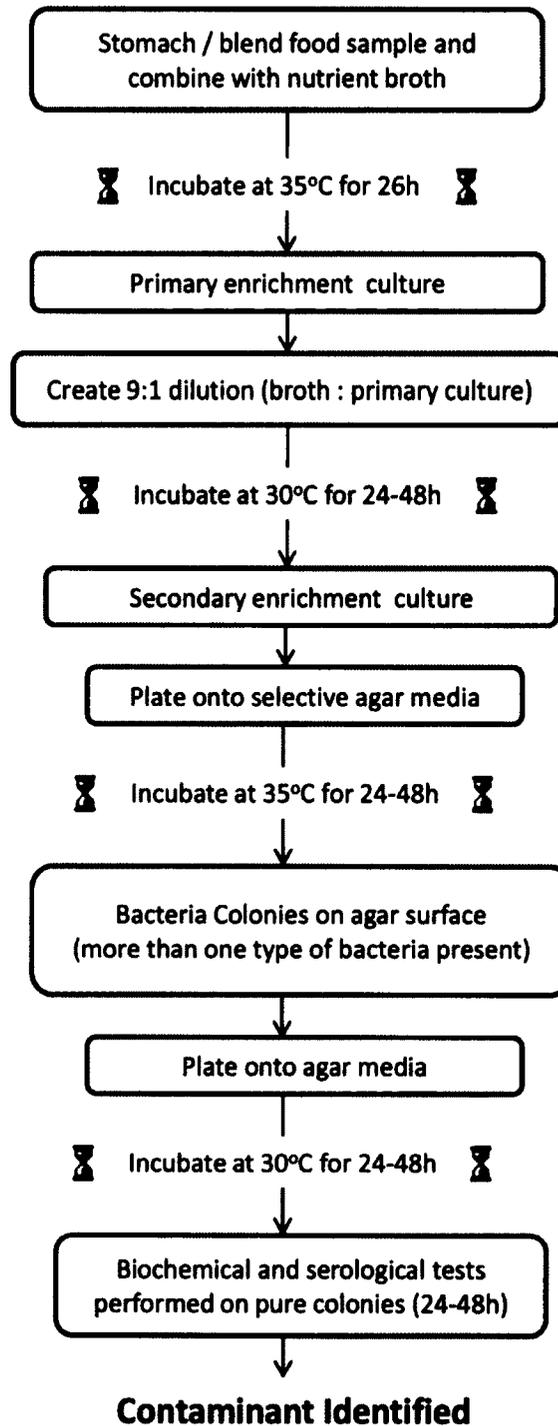


Figure 1.1: Processing stages involved in conventional methods of foodborne pathogen detection. This is a generalized view of the process, illustrating multiple culturing/incubation phases before results become available.

The disadvantages stated above pose significant problems to food inspection authorities. When a food sample is suspected of contamination, it is desirable to have reliable results available *as rapidly as possible* so that appropriate actions can be taken. For instance, if that food has already reached market and is being sold to the public, a positive pathogen result may result in a *recall*, which ensures that the adulterated product is removed from sale quickly. The longer it takes to ascertain the presence and nature of the pathogen, the more likely it becomes that consumers will eat the questionable product, and less likely that the affected food will be recovered. The result is an increased risk of higher infection rates to the public, resulting in food related illness [13]. In a food manufacturing facility, HACCP monitoring requires rapid assay results to ensure that products are free from contamination at key phases of production. Delayed screening results diminish the overall productivity of the facilities and make it more likely that a processed contaminated product will make it to market. It should come as no surprise that the high cost of conventional testing methods is problematic. Food inspection activities are generally carried out by government-funded regulatory authorities (who recently have been under constant pressure to reduce costs to control deficit spending) and industry (who are concerned with reducing these costs as much as possible to maximize profitability without compromising safety). Finally, since conventional food testing methods are generally of limited portability, it becomes difficult to establish inspection systems in places where they are currently unavailable but sorely needed, as is the case in the developing world [14].

1.3 Thesis Objectives

Since this is such a well acknowledged problem, there already exists a wide range of newer sensor technologies that have recently been introduced to improve upon one or more of these inadequacies. The most studied are a class of analytical devices that can loosely be classified as *biosensors*. Generally, a biosensor converts a response from a biological entity (such as a microorganism) into an electrical signal with some transducer mechanism (e.g., optical or electrochemical). The resulting signal can be then be amplified, digitized and further processed with a computer [10], [11], [15–17].

In this thesis, the overall objective was to research the feasibility and potential of using biosensor-based measurement systems as attractive alternatives (in terms of reduced testing time, cost and complexity) to conventional methods in food safety applications.

In undertaking this objective, the intent was *not* to develop a new sensor type to address this problem. Rather, it was hypothesized that the capabilities of off-the-shelf biosensor systems could be enhanced by introducing novel sample preparation methods, and by judiciously selecting appropriate tools in the signal processing engineer's armamentarium. It was hypothesized that with this approach, it would be possible to extract biosensor-based features that effectively described the sample under test, and use these features in pattern classification systems that could demonstrate the potential to accurately identify the nature of various food contaminants. These methods might ultimately prove to be robust to different measurement conditions, and demonstrate improvements over conventional methods of foodborne pathogen detection and

identification in several key aspects – faster results, lower cost, and reduced need for sample preparation/enrichment. The work presented in this thesis presents exploratory work towards this goal, which could ultimately be applicable to cases originating from random food sampling (e.g., as would occur in a food processing facility) as well as food samples suspected of contamination (e.g., in a food inspection lab).

The biosensor types employed in this thesis were the *electronic nose* and *Raman spectroscopy*. Briefly, electronic nose (or *e-nose*) is the term used to describe a class of instruments that measures response patterns generated by an array of gas sensors to the volatile organic compounds (VOCs) originating from a sample under test. There are a wide range of sensor modalities (e.g., conducting polymer, piezoelectric, electrochemical) on which commercially available e-noses are based [18]. Raman spectroscopy is an optical measurement technique based on the inelastic vibrational and rotational properties of the molecules in the sample under test, when excited by a laser source. The response pattern of this spectroscopic technique is a spectrum that measures the intensity of light at various wavelength shifts [19]. Both of these are described in greater detail in later chapters.

1.4 Thesis Contributions

The overall contribution of this thesis is the development of a set of novel measurement systems and sampling methods (based on e-nose and Raman biosensors) that are applied at various stages throughout the overall food inspection process. For each, the results obtained demonstrate whether the technique is appropriate for further study as a viable alternative method for identification of foodborne pathogens and spoiled

food (that might address one or more of the problems with conventional methods mentioned above).

Specific contributions are described under the headings below, each having a reference to the position in the food inspection process (see Figure 1.2) where the relevant sample type would be available.

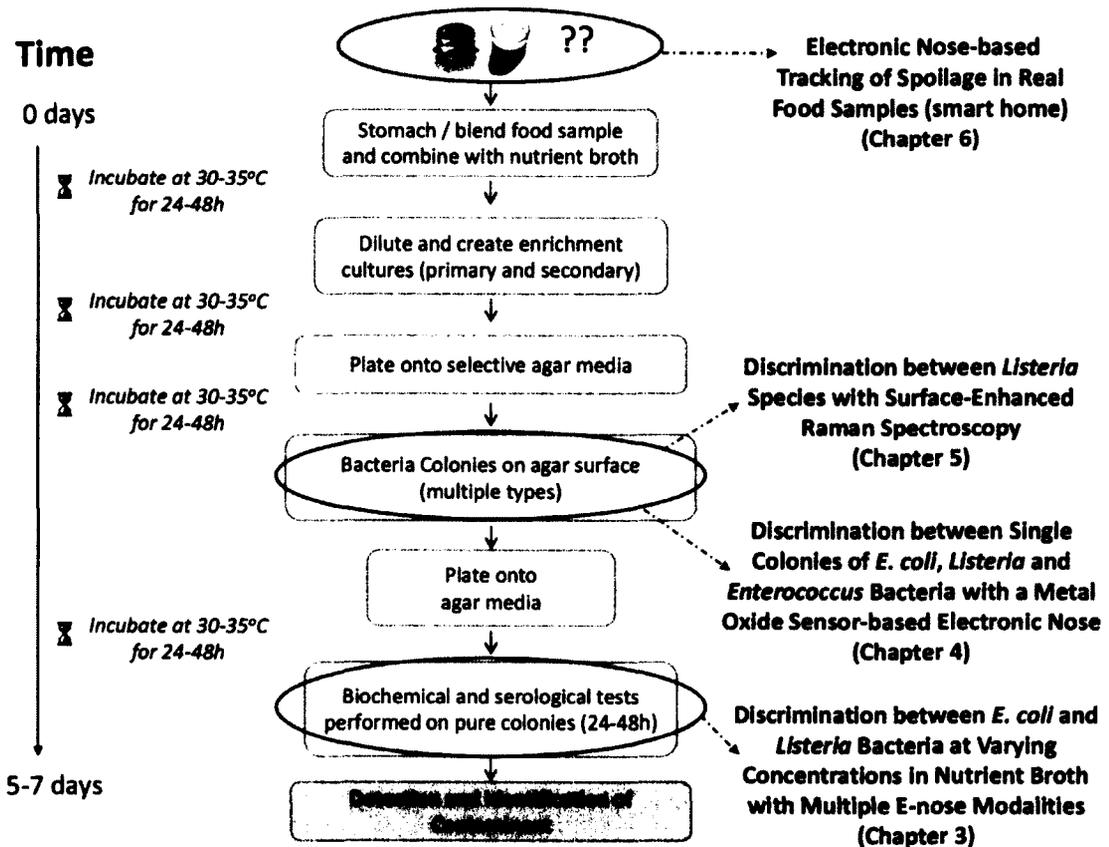


Figure 1.2: Simplified view of conventional methods, annotated with the thesis contributions and the chapters in which each contribution is described. The sample type for each contribution (and its relative position in the overall process) is indicated.

1.4.1 Contributions: Discrimination between *E. coli* and *Listeria* Bacteria at Varying Concentrations in Nutrient Broth with Multiple E-nose Modalities

An e-nose based system for bacteria discrimination was developed, and the feasibility of the e-nose to perform bacteria discrimination was demonstrated. This system was used to investigate discrimination performance at varying bacteria concentrations. For three e-

nose instruments of differing underlying modalities, discrimination between *E. coli* DH5 α and *Listeria innocua* cultures was demonstrated at high concentrations (>90% at 10⁸ cells/mL). Discrimination performance was then shown to deteriorate as the concentration was lowered (down to 10⁶ cells/mL). In doing so, it was demonstrated that the growth medium (nutrient broth) exhibited a significant time-varying volatile drift. This drift prevented the ability to assess whether the e-nose system was able to identifying bacteria at still lower concentrations. A key result of this investigation is the recommendation that performing serial dilutions of bacteria with nutrient broth is not an appropriate method for assessing e-nose discrimination potential at lower concentrations. This work represented the first time that uncorrelated linear discriminant analysis (ULDA), a supervised dimensionality reduction method, has been applied to e-nose data. For a fixed number of feature vector components, ULDA outperformed the unsupervised technique considered (principal component analysis, or PCA). Finally, preliminary data suggests that the use of ULDA (in conjunction with an altered sampling protocol designed to reduce the effect of drift in the growth medium) is able to assist with both bacteria discrimination at lower concentrations (10³ cells/mL), though further validation studies are required.

These contributions are described in Chapter 3, and the results of this work are published in [20][21]:

G. C. Green, A. D. C. Chan, and R. A. Goubran, "An investigation into the suitability of using three electronic nose instruments for the detection and discrimination of bacteria types," in *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY, 2006, pp. 1850–1853. (CONFERENCE)

G. C. Green, A. D. C. Chan, and R. A. Goubran, "Dimensionality reduction of electronic nose data for bacteria discrimination," in *Proc. 30th Conference Canadian Medical Biology Engineering Society*, Toronto, Canada, 2007. (CONFERENCE)

1.4.2 Contributions: Discrimination between Single Colonies of *E. coli*, *Listeria* and *Enterococcus* Bacteria with a Metal Oxide Sensor-based Electronic Nose

This work represents the first time that e-nose analysis of single bacterial colonies transferred from an agar plate was performed (using phosphate buffered saline (PBS) as a suspension medium). Since culture plates from food inspection labs *in general* have multiple types of bacteria present (before purification), this represents a potential improvement over other bacteria colony sampling methods, which sample the odour associated with the agar plate as a whole. The research presented herein suggests that using this novel bacteria sampling method for e-nose analysis has the potential to allow identification of bacteria on a per-colony basis. A measurement system was devised using steady-state and transient features extracted from the MOS sensor response curves with subsequent dimensionality reduction methods. Classification accuracies of over 90% were achieved when discriminating between *E. coli* DH5 α and *Listeria innocua* using this system. The success of this sampling arrangement is partially attributed to the lack of significant drift behaviour using the new PBS-based sampling method.

This initial feasibility work was then further validated with a wider range of bacteria (using the two previously considered, along with *Enterococcus faecalis* and *E. coli* Biotype I), while simultaneously controlling for growth medium and colony size (for which classification accuracies of over 85% were obtained). This work also demonstrated

that multiple consecutive e-nose responses (i.e. *sniffs*) on the same sample can be used to increase system accuracy. This is done through the use of a composite e-nose response, wherein the first sniff is omitted due to its contamination by a sensor conditioning effect.

Finally, confidence measures (based on posterior probabilities and silhouette coefficients) were calculated to assess the degree to which each sample's classifier result could be trusted. This was done to motivate this method's suitability in a pre-screening role – where the classifier output for samples with a suitably high confidence could be trusted, while other samples (with lower confidence in the classifier output) would be rejected as unclassifiable and deferred to the conventional methods.

These contributions are described in Chapter 4, and the results of this work are published in [22]:

G. C. Green, A. D. C. Chan, H. Dan, and M. Lin, "Using a metal oxide sensor (MOS)-based electronic nose for discrimination of bacteria based on individual colonies in suspension," *Sensors and Actuators B: Chemical*, vol. 152, no. 1, pp. 21-28, Feb. 2011. (JOURNAL)

1.4.3 Contributions: Discrimination between *Listeria* Species with Surface-Enhanced Raman Spectroscopy

Distinguishing different species of bacteria (*within* a genus, e.g., *Listeria*) is a more difficult problem than distinguishing bacteria of different genera (e.g., *E. coli* vs. *Listeria*). A measurement system capable of discriminating between *Listeria* species with high accuracy was developed. The system consists of portable SERS-based Raman measurement hardware (developed by CFIA) in combination with a novel spectral processing methodology (based on wavelet feature extraction and dimensionality reduction). This system resulted in classification accuracies exceeding 95% for

discriminating amongst six *Listeria* species as well as identification of the pathogenic *L. monocytogenes* species. This work represents the first time that discrimination between *Listeria* species has been reported with a SERS-based system. This exploratory work suggests that with the use of Raman spectroscopy, the overall amount of time required to identify a bacterial species can be reduced (compared to conventional methods). Classifier outputs are reported with confidence measures that could potentially be used to identify those samples for which the system's classification performance is deemed to be unreliable.

These contributions are described in Chapter 5, and the results of this work are published in [23][24]:

G. C. Green, A. D. C. Chan, B. S. Luo, H. Dan, and M. Lin, "Identification of *Listeria* Species Using a Low-Cost Surface-Enhanced Raman Scattering System With Wavelet-Based Signal Processing," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 10, pp. 3713–3722, 2009. (JOURNAL)

G. C. Green, A. D. C. Chan, R. A. Goubran, B. S. Luo, and M. Lin, "A rapid and reliable method of discriminating between *Listeria* species based on Raman spectroscopy," in *Proc. IEEE Instrumentation and Measurement Technology Conference*, Victoria, Canada, 2008, pp. 513–517. (CONFERENCE)

1.4.4 Contributions: Electronic Nose-based Tracking of Spoilage in Real Food Samples (with possible applications to Smart Home)

This work represents the first time that an exploratory analysis into the feasibility of using e-nose technology for smart home applications (as a potential assistive technology for older adults) has been performed. Using food spoilage as the odour event of interest, it is demonstrated that a commercially available e-nose can track spoilage trajectories (in

both the raw sensor space and dimensionality-reduced feature space) for multiple foods in a repeatable manner. The experiments demonstrated that the e-nose responses associated with spoilage varied depending on: a) the manner in which the sampling was performed (e.g., using the same or unique aliquots), b) the headspace generation parameters, and c) the storage temperature of the food sample prior to processing. Signal processing components used to demonstrate these effects included PCA, multi-layer perceptrons and fuzzy clustering.

More generally, measuring odours in this setting can be useful for monitoring activities of daily living (ADL) as well as identifying potentially unsafe conditions to the occupant. A novel framework for including an odour monitoring system in a smart home is proposed, key elements of which include a distributed e-nose array, wireless communication and alerting systems, integration with other expert systems, and a multi-expert fusion block.

These contributions are outlined in Chapter 6, and the results of this work are published in [25][26][27]:

G. C. Green, A. D. C. Chan, and R. A. Goubran, "Monitoring of food spoilage with electronic nose: potential applications for smart homes," in *3rd International Conference on Pervasive Computing Technologies for Healthcare*, London, UK, 2009, pp. 1–7. (CONFERENCE)

G. C. Green, A. D. C. Chan, and R. A. Goubran, "Tracking Food Spoilage in the Smart Home using Odour Monitoring," in *Proc. IEEE Symposium on Medical Measurements and Applications*, Bari, Italy, 2011. (CONFERENCE)

G. C. Green, A. D. C. Chan, and R. A. Goubran, "Identification of food spoilage in the smart home based on neural and fuzzy processing of odour sensor responses," in *Proc. 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN, 2009, pp. 2625–2628. **(CONFERENCE)**

G. C. Green, A. D. C. Chan, and R. A. Goubran, "Integration of Odour Monitoring into a Smart Home", submitted to *IEEE Transactions on Information Technology in Biomedicine*. **(SUBMITTED DECEMBER 2011)**

Chapter Two

Background and Literature Review

In this chapter, a literature review of conventional and emerging biosensor-based food inspection methods is presented, with particular attention given to highlighting the advantages and disadvantages of each. Where appropriate, relevant background information is presented for the reader. The chapter sections are organized as follows. Section 2.1 provides a review of the conventional methods, which have been in use for decades and are still regarded as the gold standard in food inspection. Section 2.2 reviews the newer technologies (emerging in the last decade or so) that, motivated by some of the limitations inherent to the conventional methods, are currently being investigated for use in this application. Section 2.3 focuses on biosensor-based methods. As mentioned in the previous chapter, this thesis focuses in particular on the use of two specific biosensor systems (e-nose and Raman spectroscopy). Section 2.4 describes in some detail the principle of operation of e-noses (with particular attention given to the instruments used in the thesis) and presents a review of the research that has been done to use these instruments in food inspection. Section 2.5 provides an analogous treatment for Raman spectroscopy.

2.1 Conventional Methods

In the previous chapter, the conventional methods were introduced and some of their important limitations were mentioned. Here, in expanding on this description, the methods of *detection* and *identification* (used to determine the presence or absence of specific organisms and identify their origin) are emphasized, while less attention is given

to methods of *enumeration* (used to quantify, for instance, the total number of bacteria in a sample) [28].

2.1.1 Enrichment, Colony Growth, and Purification

Though the procedures used for identifying different bacteria are unique for each [29–31], the following provides a general description of a process that is followed when a food is being tested for a specific target organism, adapted from [32]. A test food sample typically consists of 25g (solid) or 25mL (liquid) that is then blended or stomached to ensure homogeneity [33]. At this stage, it is very difficult to detect bacteria directly at the very low levels that can be harmful to human health (e.g., 10 colony forming units (cfu)/g for *E. coli* O157:H7). For this reason, the sample is pre-enriched by combining a portion with a non-selective nutrient broth, usually in a 9:1 (broth to sample) ratio. The mixture is left to incubate at warm temperatures (typically 35°C) for up to 24 hours, which fosters growth across a wide variety of possible organisms. This creates the *primary enrichment culture* wherein any bacteria present can reach counts of $10^2 - 10^4$ cfu/mL. Following this, a diluted portion of this culture is again combined with a *selective* broth (in a similar ratio) that preferentially supports the target organism and suppresses others (e.g., Fraser broth for *L. monocytogenes*). After further incubation at 24-48h, this *secondary* or *selective enrichment* provides sufficient growth (up to 10^6 cfu/mL) so that the bacteria can then be isolated. To do this, a small amount is plated onto one or more selective agar media and again left to incubate for 24-48h, after which distinct colonies emerge on the agar surface. For certain bacteria (e.g., *Listeria*), this process is repeated by selecting

colonies and using them to inoculate another selective agar medium (followed by a further 24h incubation period [32][34]) .

2.1.2 Confirmation Tests

At the end of the process described above, colonies have been purified and are suitable for confirmation tests to determine their nature. The types of tests used depend on the suspected organism, but they roughly fall into the following categories [35]:

1. **Morphological** – identification is based on the properties of the organism's structure and shape (e.g., rod-shaped or coccus-shaped), and
2. **Biochemical and Physiological**– identification is based on the metabolic properties (feeding and oxygen requirements) of the organism (e.g., catalase and oxidase reactions).

Applying a suitable battery of tests will differentiate between the organisms, since each will respond to the confirmation tests differently (as shown in Figure 2.1) [36]. There are kits available commercially that assist in the automating the analysis, such as the API tests (Biomérieux, Marcy-l'Étoile, France) [37].

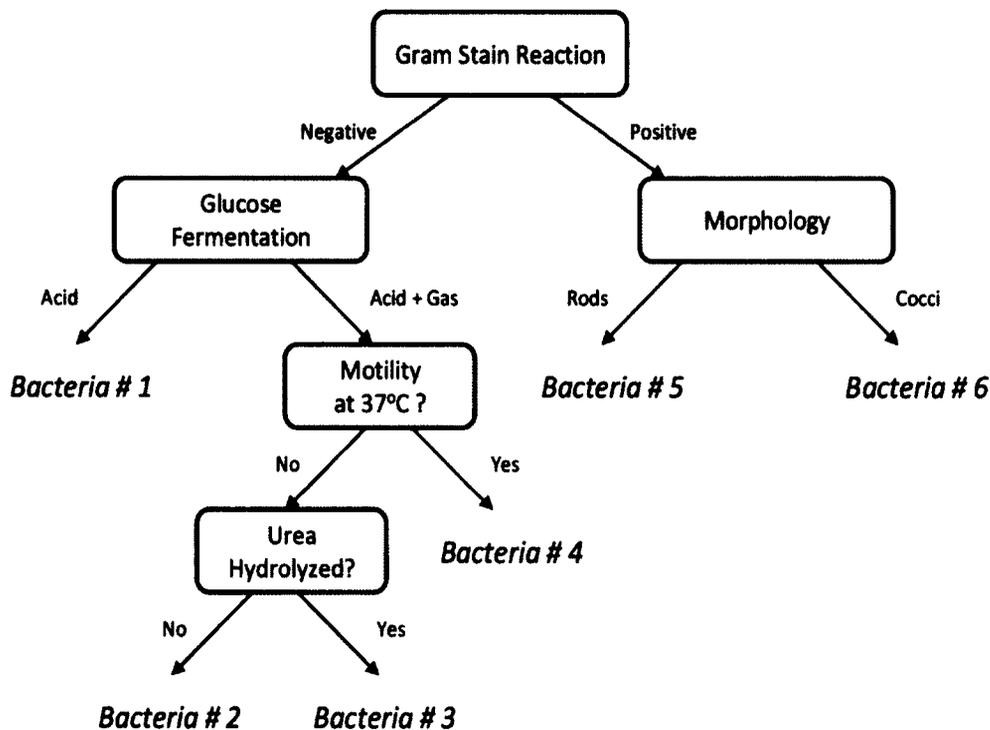


Figure 2.1: General representation of an example of a dichotomous tree used in bacteria identification. Successive tests (e.g., morphological, biochemical) continuously refine the possibilities until the result is unequivocal.

2.1.3 Advantages and Disadvantages

The classical methods are still treated by food inspection authorities as the gold standard. They remain in widespread use for several reasons:

1. they are well understood, having been developed by international consensus,
2. the methods are publicly available (not proprietary), and
3. they have demonstrated unquestionable effectiveness (both in terms of specificity and sensitivity¹) for a wide range of contaminants, including the

¹ Throughout the thesis, the terms *sensitivity* and *specificity* are used descriptively (as opposed to mathematically). Sensitivity refers to the ability of an instrument's sensors to respond discernably in the presence of a sample under test. Specificity, however, refers to the ability of that instrument to respond uniquely to different test samples.

prominent food pathogens *Salmonella*, *E. coli* O157, *Listeria monocytogenes* and *Campylobacter* [38].

As mentioned in Section 1.2, the major disadvantages of these methods are the amount of labour and time required for a categorical result. For instance, detection of *L. monocytogenes* can take up to seven days to give a result using standard methods [34][39]. Another limitation is related to bacteria that have entered the *viable but nonculturable* (VBNC) state, wherein cells enter a temporary condition characterized by a very low metabolic rate. VBNC bacteria do not undergo division and growth and thus, are not amplified by the culture and enrichment procedures described above. Having escaped detection by conventional means, these cells (which can include pathogenic bacteria such as *L. monocytogenes*, *E. coli* and *Salmonella*) are able to revive and multiply at a later time [40]. These drawbacks, coupled with the large amount of required media and reagents (all of which translates to increased cost), has spurred interest in the adoption of rapid alternative methods in this domain.

2.2 Enhancements to Conventional Methods

There exists a wide range of emerging methods (in various stages of research and product development) that seek to provide reliable results in a more expedited manner than those described above. For purposes of this review, attention is limited to two general categories: a) enhancements to conventional methods and b) biosensor-based methods. The methods that are used in this thesis – Raman spectroscopy and e-nose – fit into the latter category and the use of these instruments for pathogen detection is reviewed in greater detail at the end of this chapter.

The methods described below are considered enhancements to traditional methods inasmuch as they usually require culturing and enrichment as a pre-requisite step. There are a great number of techniques that have been proposed, but in what follows, focus is placed on the more established ones. It should also be noted that there is another class of methods that seek to expedite the results from conventional food testing. These offer improvements by degree (as opposed to an entirely new technology) and include better automation tools (in terms of agar preparation, dilution and plating), methods that aim to reduce the need for lab infrastructure (e.g., 3M Petrifilm), and one-step enrichment broths [38].

2.2.1 Immunoassays

This class of tests for bacterial presence exploits the specific manner in which antibodies and antigens behave. Antibodies are large protein molecules generated by the immune system that ward off specific contaminants (called antigens, or antibody generators). The unique shape of antibodies allows them to pair with specific antigens (for example, portions of bacteria and virus cells from a food contaminant) in a “lock-and-key” fashion [41]. In food testing, antibodies of a well-defined specificity are immobilized onto a solid substrate. *Analytes* (the substance being analyzed - in this case, either the food slurry or the enrichment dilution) are then added – the one(s) that contain the target antigen will produce the specific antigen-antibody complex. The simplest (non-quantitative) tests in this category use visible clumping (agglutination) as the antibody-antigen reactions occur as the indicator of a positive result [42].

A popular variation called *ELISA* (for enzyme-linked immunosorbent assay) uses an enzyme bound to a specific antibody. In a direct ELISA test (in which the antigen itself is detected), the antibody is deposited onto the walls of a microtiter plate. The test substance is added to the wells and if the complementary antigen is present, binding will occur. The bound antigen withstands a subsequent rinsing phase that removes the rest of the test sample. Following this, the same antibody (modified to carry a reporter enzyme) is introduced and binds to the antigen in a “sandwich” configuration. After further rinsing (removing unbound antibodies), the substrate is added, resulting in a optical change if the antibody-antigen binding has taken place (and no change otherwise), which can be detected by either fluorometric or colorimetric techniques [43].

A disadvantage of these techniques is that they are generally considered to be indicative but inconclusive, meaning that further confirmation is usually required [38]. In addition, the detection limit can be as high as 10^4 - 10^5 cfu/mL. Thus, the highly specific but low sensitivity of these methods can be unsatisfactory for certain applications.

2.2.2 Genetic Methods

With these methods, contaminants are identified by “signatures” of genetic information coded in specific deoxyribonucleic acid (DNA) sequences. In order to achieve specificity for a target microorganism, it is necessary to specify a genomic DNA segment characteristic to that organism (this is called a *target*). This task has been made easier with the wealth of publicly available DNA datasets. The most widely used genetic method is the *polymerase chain reaction* (PCR) [44][45]. This method is essentially a

procedure for *amplification* of the DNA target. It is performed by repeated iterations of the following:

- *denaturation* (heating the test sample to 94-96°C to split the DNA double helix into single strands),
- *annealing* at 50-65°C (allowing DNA primers to bind to the single strands at matching positions, which initiates synthesis of the target DNA sequence), and
- *extension* (wherein a DNA polymerase is used to catalyze the creation of the target DNA strand, if present in the test sample).

During each iteration, the amount of target DNA is doubled. Once this reaction is complete, it is necessary to determine if the reaction amplified the anticipated target – this is typically done with agarose gel electrophoresis.

Though widely used (due to its high sensitivity and ability to generate rapid results), a limitation of PCR and its derivative techniques is that it cannot be used to make a distinction between viable and non-viable cells, since the same DNA is present in both. In addition, much like the conventional methods, it can be expensive and complicated, requiring skilled operators to run the tests [10].

2.3 Biosensor Methods

During the past couple of decades, a new class of devices has emerged with potential applicability to detection and identification of foodborne pathogens. Known as *biosensors*, they differ from the methods described already in that an underlying biological response (generated by a *bioreceptor*) is coupled with a transducer to elicit an

electrical signal, which can then be amplified, digitized and processed with computers [10]. For pathogen detection, the biological response should be:

- a) *sensitive* enough to generate a measurable signal over a wide range of contaminant concentrations,
- b) *specific* enough that the response generated due to the pathogen is notably different from background activity, or activity due to other species, and
- c) *robust* to different measurement conditions (such as temperature and humidity).

The use of biosensors can potentially decrease the amount of time between sampling and the availability of results (and thus overcome the major limitation with conventional methods and their variations already described in Sections 2.1 and 2.2). Moreover, since miniaturization is possible (these devices can be fabricated with manufacturing processes borrowed from the electronics industry), it could lead to smaller sample and reagent needs (thereby decreasing operating costs) and the development of portable systems capable of *in situ* monitoring.

Recording the signals on a computer allows the use of a wide array of signal processing and pattern recognition techniques (implemented in software) that can be used to extract meaningful features and build classifier models. In reviewing the literature, it should be noted that the judicious use of *chemometrics* (that is, pattern recognition algorithms applied to chemical species) contributes significantly to the success of biosensor-based pathogen identification.

A full review and taxonomy of biosensors would elaborate upon both the bioreceptors and transducers. The bioreceptor can be any biological element (such as bacteria or metabolic byproducts, enzymes, proteins, DNA, antibodies) that interacts with the

transducer. The review below is restricted to transducer technologies and their use in the foodborne pathogen detection problem, since this is the primary area of interest.

2.3.1 Biosensors based on Optical Properties

For optical biosensors, the bioreceptor interaction results in a *change in the properties of light* (either reflected or incident), which is then measured by the transducer to generate a spectrum. Spectroscopy is frequently used in identification or “fingerprinting” applications, with the underlying principle that the measured spectrum is a unique property (or “signature”) of the substance under test.

One of the best known techniques is *Fourier-transform infrared spectroscopy (FTIR)* which measures the amount of incident light (as a function of wavelength) that gets absorbed by a sample [46]. Absorption varies with wavelength because certain frequencies match the resonant frequencies of the covalent chemical bond vibrational modes (e.g., stretching, scissoring) in the sample’s molecular structure. The ability to identify and differentiate bacteria based on FTIR was first demonstrated by Naumann et al. [47]. Various genera of bacteria were incubated, suspended in water, transferred to an optical window, and dried to form a transparent bacterial film. The generated spectra (between 500-4000 wavenumbers) were measured. Subsequent calculation of the spectral second derivative, followed by correspondence map calculations led to a clear separation between types. Since this work, a large literature has emerged demonstrating similar results using a whole host of bacteria types and pattern classification techniques. For example, in [48] discrimination is reported at the species level when the spectral features are combined with an artificial neural network (ANN) architecture employing

hierarchical cluster analysis (HCA) as a pre-processing step. The use of FTIR requires that an initial culturing stage be performed, but after that sample preparation is relatively simple and results are rapidly available.

Other optical biosensors in addition to FTIR are based on:

- *surface plasmon resonance* (SPR). This technique is based on measuring the changes in the *refractive index* of incident light, due to cell bindings with the transducer surface. The ability of SPR to discriminate pathogens is described in [49],
- *optical fibers*. Though initially developed for telecommunication, biosensors are now based on optical fibers that exploit the phenomenon of total internal reflectance [50], and
- *Raman spectroscopy* (of particular interest in this thesis, and described further in Section 2.5).

2.3.2 Biosensors based on Mass / Acoustic Wave Properties

This class of sensors uses a piezoelectric crystal (typically quartz) that exhibits a stable oscillation in the presence of an applied voltage in conjunction with a sensor coating. When the crystal interacts with a biological sample (due to surface binding chemicals), the mass changes slightly (*mass-loading* effect). This effect alters the oscillation frequency of the crystal, and this property is measured to characterize the sample [10]. One configuration, the quartz crystal microbalance (QCM), was used by Su and Li [51] to detect *Salmonella Typhimurium*. *Salmonella* antibodies were used to coat the crystal (resonant frequency of 7.995MHz) and an analyzer measured the changes in

resonant frequency with subsequent exposure to antibody and antigen in varying concentrations. In [52], Berkenpas et al. used a surface acoustic wave (SAW) device (in which the piezoelectric material is used to generate the acoustic waves), again with antibody in the sensing region, in order to detect *E. coli* O157:H7. Modern circuit technology makes fabrication of these biosensors very straightforward and cost-effective.

2.3.3 Biosensors based on Electrochemical Properties

In contrast to the above, electrochemical biosensors exploit the fact that due to their underlying chemistry, biological samples exhibit changes in their electrical properties (e.g., ion transfer, resistance) that can be measured by a transducer and used to characterize the sample. They have been explored extensively in the area of pathogen detection, as reviewed in [10]. The most common methods in this category are *amperometric* (wherein the sensor applies a potential difference that results in electrical current as the measured property). In most cases, the current is measured not directly from the food sample but in the presence of an immunoassay (as described above). For example, Chemburu et al. [53] were able to use dispersed carbon electrodes in combination with a flow-through ELISA to measure currents (in the microamp range) that correlated with varying concentrations (50-1000 cells/mL) of *Campylobacter jejuni* and *L. monocytogenes*. Muhammad-Tahir and Alocilja [54] used a disposable conducting polymer-based sensor (also in the presence of antibodies) to quantify the degree of *E. coli* O157:H7 contamination in spiked fresh produce (strawberries, lettuce and alfalfa). Other electrochemical biosensors are categorized as potentiometric or impedimetric, based on the electrical property measured (potential and impedance, respectively). Electrochemical

biosensors have the advantages of low cost, potential miniaturization, and fast assay times. In general, however, the sensitivity and specificity are limited when compared to other methods.

The biosensor types above form the basis of several categories of e-nose instruments, which are described next.

2.4 Electronic Nose (E-nose) Biosensors

2.4.1 Principle of Operation

Over the last couple of decades, the use of e-nose technology has become an attractive alternative to some of the traditional analytical methods described previously.

As defined by Gardner and Bartlett in 1994, an e-nose is:

“...an instrument, which comprises of an array of electronic-chemical sensors with partial specificity and an appropriate pattern recognition system, capable of recognizing simple or complex odours [55]”.

Electronic noses have the following properties [18]:

- the sensor(s) respond in some manner to the odorant molecules (VOCs) emanating from the gas phase (or *headspace*) of a sample under test. Several e-nose sensor modalities are available, each with an underlying physical phenomenon that gives rise to the sensor response (e.g., conductivity changes or mass/acoustic wave properties),
- the sensor(s) are arranged in an *array* configuration, either physically (the array consists of multiple physical sensors) or logically (the array is a set of

measurements derived from a single physical sensor), or a combination of both, and

- each sensor in the array has a broad range of selectivity (i.e. it will respond to a wide range of samples) but the sensitivity of each of the sensors varies for different samples. This provides a pattern of sensor responses, called a *smellprint* or *sniff*, which is unique for every sample in a wide range of substances. This is in contrast to substance-specific detectors (e.g., carbon monoxide alarms) which are tuned to respond to the presence of a single chemical only.

In commercial systems, this gas sensor array is packaged with sample handling and data processing modules. With appropriate *training*, the e-nose can learn the smellprints associated with certain categories of samples and identify the nature of an unseen test sample. This sensor system mimics some of the biological processes that occur in both nasal and neural pathways used in the human sense of smell (or olfaction). For this reason, the terms electronic nose, *artificial olfaction*, and *machine olfaction* have become popular in the scientific literature. It is important to realize, though, that an e-nose instrument is not designed to be an exact analog of the human olfaction system, which is incredibly complex [56]. It should also be emphasized that e-nose analysis is fundamentally different from traditional analytical techniques, such as gas chromatography and mass spectrometry (GC/MS). Whereas GC/MS identifies the *origin* of the components in a complex mixture, the e-nose does not – exactly why the sensors interact with the sample headspace in the way that they do is generally not well understood. In this sense, e-noses are empirical instruments that use the differences in sensor response patterns (rather than any absolute chemical or physical property) to

discriminate between sample types. E-nose technology has been widely commercialized and is now deployed for diverse applications (e.g., quality control in the pharmaceutical, packaging and food industries). Figure 2.2 shows the main components of an e-nose system.

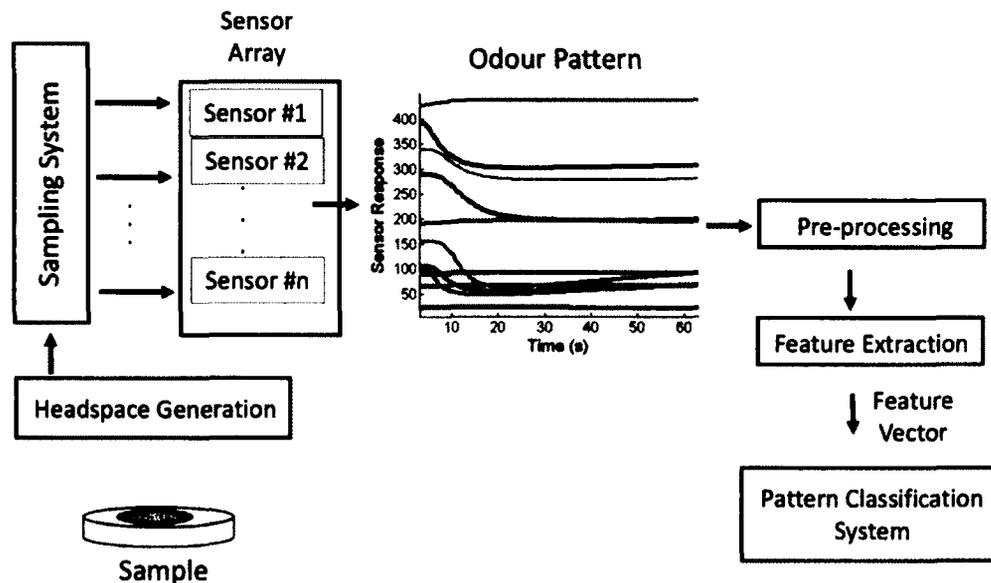


Figure 2.2: Main functional components in an e-nose measurement system.

With this background, the major advantages of using e-noses for detection of foodborne pathogens can be appreciated [57]:

1. It is rapid – once trained, an e-nose can provide classification results very quickly depending on sensor type (typically minutes),
2. Sample preparation is straightforward,
3. It is relatively inexpensive (as compared to analytical techniques such as GC/MS),
4. E-nose technology is increasingly being commercialized in portable configurations, such as the Cyranose 320 (Smiths Detection, Pasadena, CA, USA) [58] and the zNose (Electronic Sensor Technology, Newbury Park, CA, USA) [59]), and

5. Specific biochemical reagents are not required.

2.4.2 Electronic Nose for Identification of Bacteria

The potential to use e-nose technologies for microbial detection and identification has been investigated for over a decade with promising results, and the application domains include medical diagnosis, foodborne pathogen detection and bio-terrorism [60–62]. Intuitively, this is not surprising, since it is common for a trained microbiologist to be able to identify some species based on the *smell* of a culture [63]. It has also been reported that in certain biomedical applications (e.g., detection of cancer), other mammals (particularly dogs) have the ability to be trained to recognize the VOCs associated with the pathology of interest from smelling various types of complex samples (e.g., urine or breath) [64]. While canine olfaction may be unparalleled in its sensitivity, there are advantages that an e-nose based approach offers when compared to mammals – objective results and reduced costs for training are principal among these. Further enhancements to e-nose sensor technology (increased number of sensors) may further increase the sensitivity of these methods while preserving these advantages.

Using a 16 sensor CPS e-nose combined with a multilayer perceptron (MLP) neural network, Gibson et al. obtained an overall classification accuracy of 93.4% when discriminating between 12 bacteria and one yeast, though the bacterial concentrations are not reported [63]. Gardner et al. describe an e-nose (based on 6 metal oxide sensors) developed to distinguish *E. coli* and *Staphylococcus aureus* (in nutrient broth), as well as identify its growth phase (lag, log, or stationary). They used an optical densitometer to measure the bacteria concentrations (which ranged roughly from 10^7 – 10^9 cells/mL) during their experiment. Using a sophisticated MLP classifier with cross-validation, the

best classification results reported (after investigating simultaneously a wide range of features and normalization methods) was 96.1% for bacteria type and 80.7% for growth phase. It should be noted that the classification used measurements over the full range of concentrations, since the data were sampled regularly during all phases of growth [65]. Holmberg et al. obtained bacteria from urinary tract infection patients (subsequently grown on agar plates). With an e-nose of 15 gas sensors (of multiple modalities) and a feature extraction stage using the parameters of fitted exponential curves, accuracies of 68-76% (depending on the classifier) were reported for discriminating between five bacteria classes. Interestingly, they found that the signals obtained from concentrations corresponding to lag growth phase (which were not provided in absolute terms) were insufficient to detect bacteria, and suggest that this is due to the predominance of the volatile activity associated with the agar medium in the early stages [66]. Using genetic algorithms to optimize a backpropagation MLP, Pavlou et al. were able to distinguish between five types of gastroesophageal bacteria at concentrations of 10^7 cells/mL (in both pure and mixed cultures). They achieved 98% predictive accuracy (for discrimination between bacteria types) and 94% (for detection of bacteria in various growth media) [67]. Dutta et al. used a handheld instrument (32 conducting polymer sensors) in combination with sophisticated post-processing (using neural networks with self-organizing maps (SOMs) and fuzzy clustering in their classifier) in order to achieve 96% classification accuracy when discriminating between six bacteria types responsible for eye infections. Their samples (prepared in blood agar) were diluted to concentrations of 10^8 , 10^5 and 10^4 cells/mL, but they do not report the performance of their algorithm separately for each case [68]. In another study, Alocilja et al. prepared dilutions of pathogenic vs. non-

pathogenic *E. coli* in nutrient broth from 10^2 - 10^7 cells/mL but only reported on the ability to distinguish these in the middle of this range (10^5 cells/mL) [69].

Turning attention to bacterial contamination in foods, Balasubramanian et al. were able to differentiate homogenized beef samples that had been inoculated with *Salmonella* (at a concentration level of 0.7-2.6 \log_{10} cfu/g). Their method made use of discriminant analysis to generate models and bootstrapping to verify their performance [70]. Arnold and Senter were able to distinguish between a wide variety of bacterial genera (including *Listeria*, *Enterobacter*, and *Klebsiella*) using samples originating from both lab cultures and poultry carcasses [71]. Siripatrawan determined that *E. coli* contamination in a packaged fresh vegetable product (alfalfa sprouts) could be identified with a MOS based e-nose and a SOM pattern analysis method [72]. Labreche et al. measured e-nose responses of aging milk over several weeks and correlated the results with microbial growth curves [73]. Capone et al. [74] demonstrated that for a single substance (milk), the sensor responses traced out a smooth trajectory in the PCA space as it aged (See Section 3.2).

In the remainder of this thesis (Chapters 3, 4, and 6), research using e-nose based measurements systems are described that use: a) metal oxide semiconductors (primarily), b) fingerprint mass spectrometry and c) conducting polymer sensors. A detailed description of these sensor types is deferred to the next chapter. Finally, it should be emphasized that the e-nose constitutes a biosensor “special case” in which there are *multiple* sensors (transducers) used – it is the *pattern of response* across these sensors that is meaningful.

2.5 Raman Spectroscopy

2.5.1 Principle of Operation

Raman spectroscopy is a spectroscopic technique based on the *inelastic vibrational and rotational properties* of the molecules of a test substance. A monochromatic light source (laser) of wavelength λ_i is used to illuminate the test substance. When this occurs, laser photons interact with the electron clouds involved in the substance's molecular bonds, causing some electrons to become excited and either (as illustrated in Figure 2.3):

1. shift from an initial ground state to a virtual energy state, then relax to a vibrational excited state (this is called *inelastic Stokes Raman scattering*),
2. shift from a initial vibrational excited state to a virtual energy state, then relax to a ground state (this is called *inelastic anti-Stokes Raman scattering*), or
3. shift from an initial ground state to a virtual energy state, then relax back to the original ground state (this is called *elastic Rayleigh scattering*)

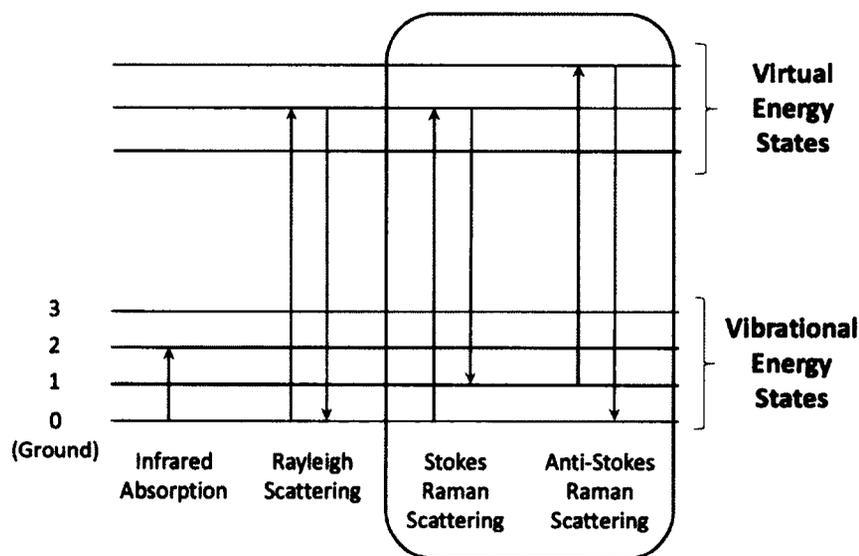


Figure 2.3: Illustration of energy levels associated with Raman spectroscopy (highlighted box).

In each case, a photon is emitted in the process (the photon's wavelength λ_e determined from the energy difference E between states using the equation:

$$E = \frac{hc}{\lambda_e} \quad (2.1)$$

where c is the speed of light (3×10^8 m/s) and h is Planck's constant (6.626×10^{-34} Js). The first two (inelastic) cases above describe the *Raman effect*, whereby the emitted photons have a wavelength different from the incident laser photons [75]. In the parlance of Raman spectroscopy, radiation is more typically expressed in terms of *wavenumbers*, ω instead of wavelength λ . These are related by $\omega = 1 / \lambda$ (wavenumbers have units of m^{-1}). The relationship between the incident and emitted light in Raman scattering is given by:

$$\Delta\omega = \frac{1}{\lambda_i} - \frac{1}{\lambda_e} \quad (2.2)$$

where $\Delta\omega$ is termed the *Raman shift* [75].

This effect is the one of interest in Raman spectroscopy measurements and is usually very weak, being exhibited in roughly only 1 in 10^7 incident photons. The last effect above, whereby the emitted photons have the same wavelength as the incident laser, is the dominant one, making it necessary to remove these photons with an optical filter. The resulting light (consisting of "Raman effect" photons) is then sent to a spectrometer, which measures the intensity of light as a function of Raman shift. This is called a Raman spectrum, and is denoted $R(\Delta\omega)$. For most biological samples, Raman shifts occur in the range $400\text{cm}^{-1} < \Delta\omega < 1600\text{cm}^{-1}$. A block diagram of the measurement system is shown in Figure 2.4. Raman spectra are contaminated with a significant

background fluorescence that, if left unprocessed, prevents the ability to resolve important peaks [76].

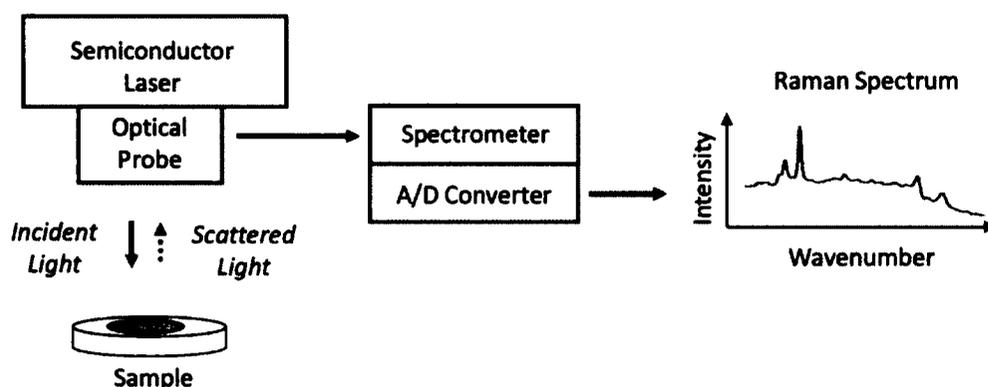


Figure 2.4: Main functional components in a Raman spectroscopy measurement system.

2.5.2 Raman Spectroscopy for Identification of Bacteria

The use of Raman spectroscopy in classifying bacteria was first reported at the cell level (for *Rhodopseudomonas* and *Sarcina*) using a variation called ultraviolet resonance Raman spectroscopy (UVRR). UVRR uses incident light in the ultraviolet portion of the spectrum (thereby reducing fluorescence) [77]. UVRR being quite expensive, further research attempted to extend these initial promising findings using the more accessible near-infrared (NIR) Raman systems [78]. In [79], the authors demonstrated that bacteria (*Staphylococcus* and *E. coli*) could be identified during early colony growth (6h), which represented strong evidentiary support for the use of Raman spectroscopy as a rapid microbiological fingerprinting method. This work also used HCA to generate dendrograms. With this and most subsequent work in this area, the use of sophisticated machine learning and pattern classification techniques figures prominently.

The introduction of surface-enhanced Raman spectroscopy (SERS) since the early work in this field has generated increasing momentum. SERS is a technique used to enhance (by several orders of magnitude) the inherently weak Raman effect, providing more usable spectra. With SERS, the test substance is combined with a mixture containing metal (typically silver) nanoparticles prior to analysis. The SERS method is described in more detail in Chapter 4.

The use of Raman spectroscopy (and in particular, SERS) holds many attractive features for detection of foodborne pathogens [78]:

1. It is non-destructive, since the laser light is typically of low intensity,
2. It is rapid – once samples and colloidal mixture are available, spectra can be obtained in minutes and importantly, the need for cell culturing is reduced,
3. It is inexpensive, from an operational point of view,
4. Accurate Raman spectroscopy systems can be constructed from portable, off-the-shelf components suitable for field deployment, and
5. Specific biochemical reagents are not required.

Compared to FTIR (the other prominent optical biosensor technology described in Section 2.3), Raman spectroscopy provides complementary information and offers several advantages [78]:

1. the ability to see spectral peaks arising from symmetric vibrations (FTIR uses asymmetric vibrations),
2. resolution improvements, both spatially (due to smaller incident light wavelengths) and spectrally (Raman spectra have peaks of narrower width, potentially allowing more detailed bacterial fingerprinting), and

3. since water does not produce a significant artifact in Raman spectra, it is suitable for aqueous samples (such as bacterial samples that arise in the natural environment).

In Chapter 5, a portable Raman measurement system capable of discriminating *Listeria* at the species level is described.

2.6 Summary and Conclusions

This chapter initially focused on reviewing conventional methods for food pathogen identification and the problems inherent in the process. Improvements to these methods were also discussed before turning attention to biosensors. A wide range of these technologies were reviewed (along with relevant research demonstrating their potential in this area). Finally, the biosensor systems used in this thesis – e-nose and Raman spectroscopy – were described and their use in foodborne pathogen detection was reviewed in greater detail.

Chapter Three

Discrimination between *E. coli* and *Listeria* Bacteria at Varying Concentrations in Nutrient Broth with Multiple E-nose Modalities

This chapter describes research that was performed to investigate the feasibility of using e-nose technology to discriminate between bacteria samples suspended in a nutrient broth. Three different e-nose modalities are used (metal-oxide semiconductors, conducting polymers and mass spectrometry), and the effect of bacteria concentration on discrimination ability is examined. In this chapter, several concepts that are central throughout the remainder of the thesis are introduced. To assist the reader, background information on these topics is presented (such as e-nose instruments, feature extraction, dimensionality reduction and pattern classification).

The work described in this chapter has been disseminated in [20] and [21].

3.1 Background

In food safety applications, high sensitivity of a detection method is desirable. For example, bacteria such as *E. coli* O157:H7 can cause infection at a concentration of only 10 cells/mL [69]. Underrepresented in the literature have been detailed investigations into the e-nose detection and discrimination limits (in terms of bacterial concentration) – most of the work reviewed in Section 2.4 is performed on highly dense cultures, and absolute concentrations are not always reported. This was recognized by McEntegart *et al.* who grew *Enterobacter aerogenes* and an *E. coli* strain in a brain heart infusion (BHI) broth and found that only after 4 hours of growth (corresponding to concentrations of 5×10^8 cells/mL) were detection and separation possible. Their analysis was performed using a

hybrid 20 sensor array (consisting of metal oxide semiconductor, quartz microbalance and electrochemical sensors) combined with simple PCA clustering. They state that “to be useful for practical applications...the e-nose approach must be greatly improved in sensitivity” [80].

The three experiments described in this chapter focused on factors important when assessing the ability to use e-nose for identification of foodborne pathogens: a) the choice of e-nose sensor modality and b) the effect of varying sample concentration on discrimination performance. Experiment #1 represented the first attempts to process bacteria samples with the three e-nose instruments available, each with a different sensor modality (MOS, FMS and CPS). An assessment of the discrimination potential of these e-nose instruments was performed with dense cultures. In Experiment #2, the ability of MOS- and FMS-based instruments to discriminate bacteria from less dense cultures (as would be available earlier in the culturing process) was assessed. Experiment #3 extended this investigation using still lower concentrations, while simultaneously adjusting the sample preparation method to attempt to mitigate the effects of a time-varying drift revealed in Experiment #2 (described below).

The work in this chapter is organized by first describing the methods used that are common to all experiments (Section 3.2). The subsequent sections provide methodology details unique to each of the three experiments (mostly related to sample preparation and sampling protocol) with the results and discussion of each (Experiment #1 – Section 3.3, Experiment #2 – Section 3.4, Experiment #3 – Section 3.5). Presenting the work in this manner highlights the motivation of adjusting the methods in each subsequent experiment

(as described below). Section 3.6 presents an overall summary and conclusions of this chapter.

3.2 Methods

The information presented in this section pertains to all three experiments – including details on the bacteria samples, e-nose instruments, pre-processing and feature vector extraction, dimensionality reduction and classification.

3.2.1 Bacteria Samples

The work described in this chapter used non-pathogenic bacteria species (*E. coli* DH5 α and *Listeria innocua*) obtained from CFIA. Both bacteria were cultured in Luria-Bertani (LB) broth at 37°C for 16-18h and then maintained at 4°C prior to testing to arrest further bacterial growth. LB broth is a nutrient-rich medium popular among bacteriologists because it provides rapid growth among a wide variety of species [81].

3.2.2 Electronic Nose Instruments Used

3.2.2.1 Metal Oxide Sensor (MOS)-based Electronic Nose

This is the most widely available type of commercial e-nose. In this instrument, the sensors are constructed from a metal oxide semiconductor (typically based on tin) that has been doped with a noble metal catalyst (e.g., platinum) and operate at high temperatures (up to 400°C). When the VOCs from the test substance are introduced, the conductivity of the sensors change, resulting in a measurable electrical signal on a pair of

metal electrodes connected to the sensor. Varying the selectivity of each sensor in the array can be achieved by changing the sensor/doping materials and/or altering the operating temperature.

In this thesis, the AlphaMOS FOX 3000 (AlphaMOS, Toulouse, France) [82] was used (see Figure 3.4(a)). This configuration used 12 sensors arranged in a two-chamber configuration, in which the first chamber contains six “LY”-type sensors (based on Cr_2O_3 and Ti) and the second chamber contains six “P and T”-type sensors (based on SnO_2 and Pt / Pd) [82]. Certain sensors (the P-, T-, and LY2/LG sensors) consist of an *n*-type semiconductor, and the resistance of these sensors will decrease when exposed to the sample headspace. Conversely, the remaining sensors consist of a *p*-type semiconductor, and the resistance of these sensors will increase when exposed to the sample headspace. The twelve MOS sensors used in this instrument were not selected for optimal performance with bacteria or food samples. Instead, the configuration was provided by the vendor for general e-nose applications. Zero air (ambient air filtered to contain less than 5ppm (parts per million) water, less than 5ppm hydrocarbons and less than 5ppm carbon dioxide) is used as the carrier gas for this MOS e-nose. Figure 3.1 shows the sensor response time curves for a MOS-based electronic nose during an exposure to a test sample. In this thesis, all MOS response curves were recorded for 120s with a sampling interval of 0.5s.

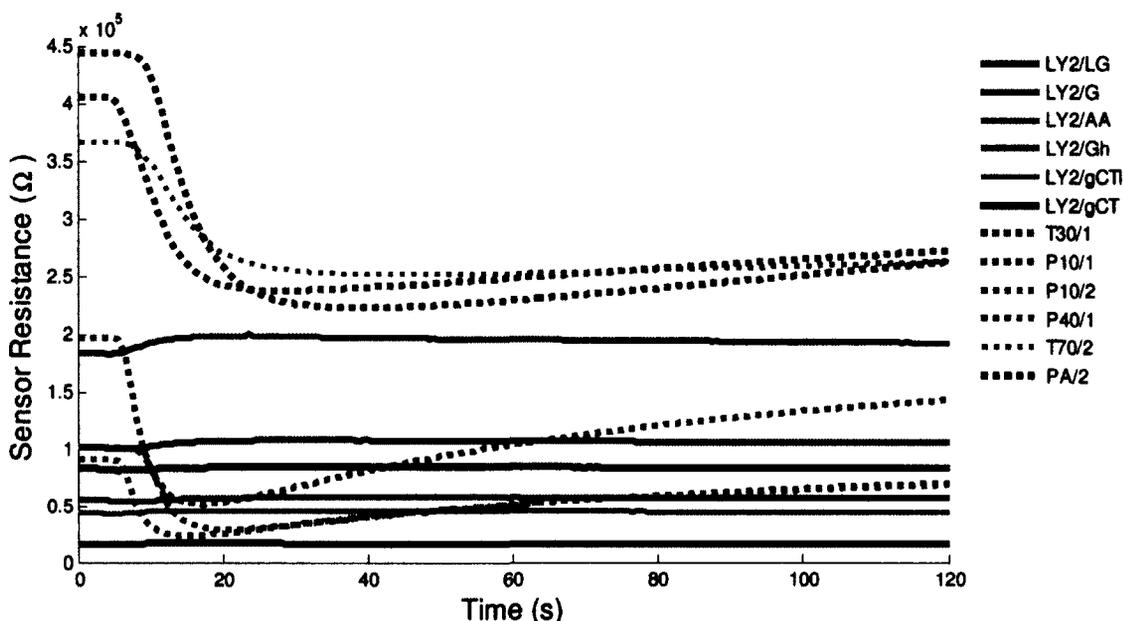


Figure 3.1: A set of sensor response time curves (before processing) for a MOS-based electronic nose (AlphaMOS FOX 3000). There is one curve for each of the 12 sensors that comprise the sensor array.

3.2.2.2 FMS-based Electronic Nose

This type of e-nose incorporates a quadrupole-based mass spectrometer to measure the mass spectrum of the test sample [83]. An ion source ionizes the sample's headspace molecules, causing the ions to travel down the length of a quadrupole (a set of 4 parallel metal rods). Varying the voltage applied to these rods affects the electric field in the ion channel, which in turn affects the trajectory of ions. For a given field strength, only ions of a specific mass-to-charge ratio (m/z) will emerge from the channel. In this way, the quadrupole acts as a mass-selective filter. The detector integrates the ion activity over time, generating a response curve of ion signal intensity vs. time for each m/z value (typically FMS systems operate from $45 \leq m/z \leq 200$). Feature extraction methods (such as integration under the time curve) are then used to derive the mass spectrum of the test sample. FMS e-nose systems differ from analytical mass spectrometers in that no separation stage is performed beforehand. The sample is characterized by its mass

spectrum *as a whole* (and not by its constituent parts) and this can be used as a fingerprint for the sample. FMS-based e-noses present an additional advantage – when the responding mass fragments are compared with existing libraries (e.g., National Institute of Standards and Technology Mass Spectral Library), it can provide additional insight into the origin of the VOCs associated with the sample.

In this thesis, the AlphaMOS Kronos was used [82] (see Figure 3.4(a)). Pure nitrogen is used as the carrier gas for the Kronos e-nose. Figure 3.2 shows the response time curves for a FMS-based electronic nose during an exposure to a test sample. All FMS response curves were recorded for 120s (representing an integration of 80 scans, each taking 1.5s).

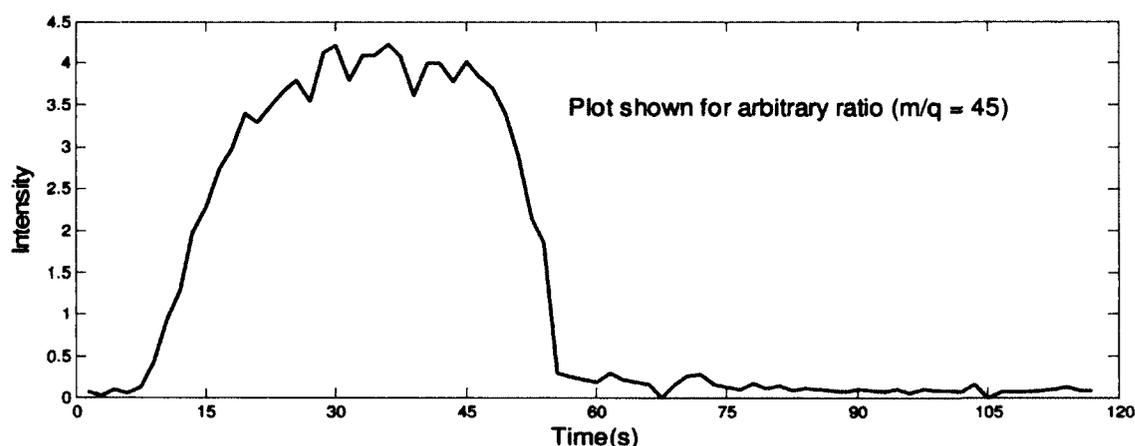


Figure 3.2: FMS response curve for an arbitrary *E. coli* sample. This curve is recorded for all m/q values – a value for $m/q = 45$ is shown. The intensity axis is in arbitrary units.

3.2.2.3 Conducting Polymer Sensor (CPS)-based Electronic Nose

Similar to MOS-based e-noses, conducting polymer sensor (CPS) e-noses are based on conductivity changes throughout a sensor array. Here, however, the active sensor material is a conducting polymer (e.g., from thiophene or polypyrrole families). When exposed to a sample under test, VOCs bond with the polymer backbone causing a

swelling effect which results in a resistance increase because the conductive pathways through the material are disrupted. When the analyte is removed, the sensor shrinks to its original size, restoring the conductive pathways. The resulting change in conductivity is measured and recorded, forming the sensor response curves. Each member of the sensor array is doped differently to provide the broad but partially overlapping sensitivities characteristic to e-noses. Unlike MOS, CPS-based e-noses operate at room temperature and thus do not require heating elements. This makes them more suitable as portable instruments [84]. It is anticipated that the cost and size of these CPS-based systems will decrease significantly in the next several years [85]. Batch variations, significant sensor drift and sensitivity to water vapour (preventing their use in variably humid conditions) are the main drawbacks associated with this type of e-nose.

In this thesis, the Cyranose 320 (Smiths Detection, Pasadena, USA) [58] was used (see Figure 3.4(b)). The Cyranose 320 is a handheld device (based on 32 CPS sensors), with on-device sampling, data processing and display capabilities². Prior to sampling, room air is drawn in through a vacuum pump to establish a baseline resistance measurement. During the “sniff” phase, the sample’s vapours are drawn through a sample inlet, combined with the room air (carrier gas) and introduced across the sensor array. The use of ambient room air as carrier gas simplifies the instrument (as compared to the MOS and FMS), making it capable of portable odour sampling. However, the lack of an absolute reference control makes the processing more susceptible to variations in the ambient environments in which it is used, which may complicate the data analysis. Figure 3.3 shows the response time curves for a CPS-based electronic nose during an exposure

² It should be noted that since performing this experiment, the Cyranose 320 has apparently been discontinued (there is no longer any mention of the product on the Smiths Detection website).

to a test sample. All CPS response curves were recorded for 120s with a sampling interval of 0.5s.

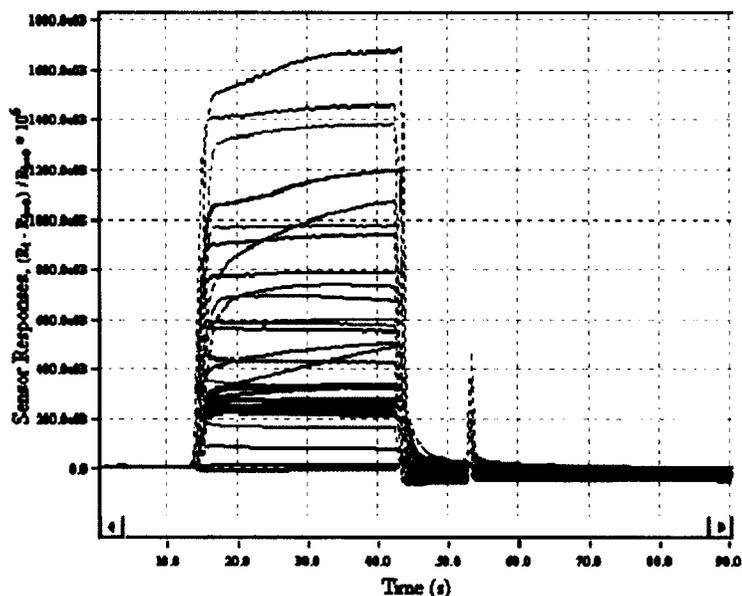


Figure 3.3: A representative set of sensor response time curves (unprocessed) for a CPS-based electronic nose (Smiths Detection Cyranose 320). There is one curve for each of the 32 sensors that comprise the sensor array. The sampling inlet connects the sampling pathway to the sensors at $t=12$ s and the purge cycle (room air) begins at $t=42$ s.

3.2.2.4 Autosampler

Both AlphaMOS instruments (the MOS- and FMS-based e-noses) are housed in the same physical enclosure along with the AlphaMOS HS100 [82], an autosampling system that allows precise control over the headspace generation and sampling process (see Figure 3.4(a)). Prior to odour measurement, aliquots of the sample of interest are transferred to glass vials (typically 10mL) and capped, an arrangement that is able to concentrate headspace in the top portion of the vial (above the sample). The HS100 consists of a robotic arm that operates a syringe mechanism capable of: a) moving a vial to and from an integrated oven/agitator used to generate headspace, and b) injecting headspace collected from the vial into the MOS and FMS instruments. Sampling is done

from up to two sample trays (with optional temperature control), each of which can hold a maximum of 32 vials. During sampling, the syringe temperature is maintained at 5°C above the headspace temperature in order to avoid condensation in the syringe. In contrast, the handheld Cyranose 320 (the CPS-based e-nose) is designed to sample VOCs in ambient air – thus, it does not have a method to precisely control headspace generation parameters. The headspace generation parameters (time and temperature) varied for each experiment, and thus will appear in subsequent sections.

Figure 3.4 shows a photograph of all instruments used.

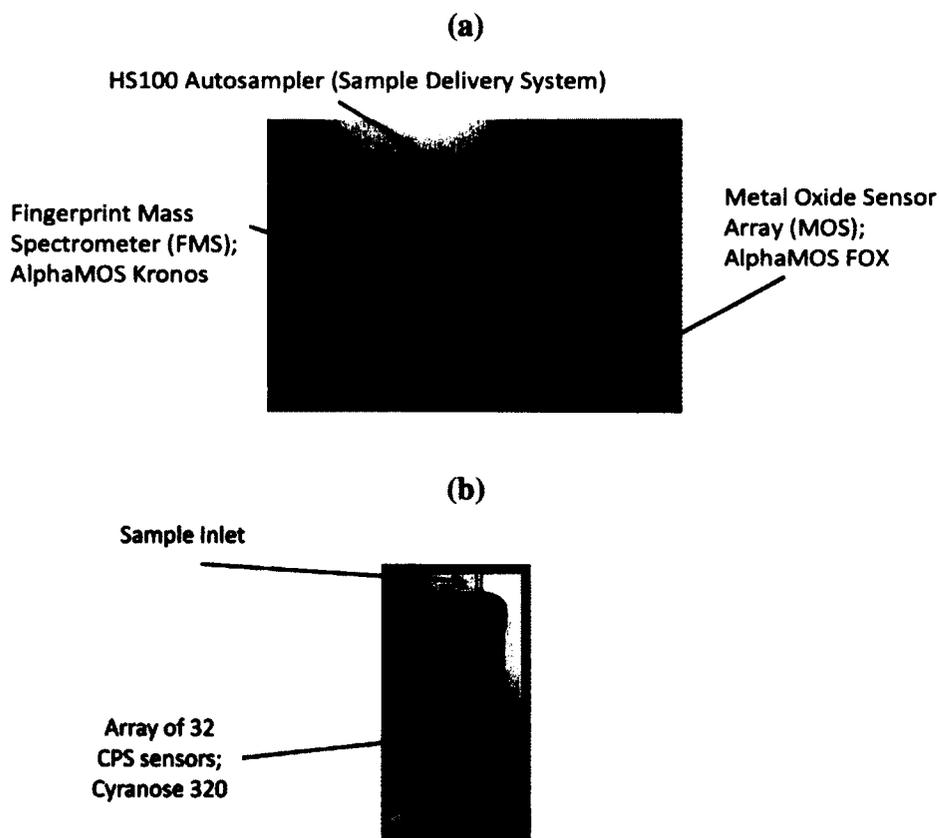


Figure 3.4 (a) A photograph illustrating the components used in the AlphaMOS e-noses – FMS and MOS. The HS100 autosampler is also shown. (b) A photograph illustrating the components used in the Smiths Detection Cyranose 320 CPS-based e-nose.

3.2.3 Pre-processing and Feature Extraction

3.2.3.1 Baseline Manipulation

In e-nose systems, baseline manipulation is typically performed to eliminate the effect of differing absolute baseline values (i.e. each sensor's response in the absence of any sample) from the measurement, thus increasing contrast and removing the effect of short-term baseline drift [86]. With the MOS and CPS, a *fractional* baseline transformation (typical in the e-nose literature) was performed on the raw sensor curves:

$$r_{j,preproc}(t) = \frac{R_j(t) - R_{j0}}{R_{j0}} \quad (3.1)$$

where $R_j(t)$ is the measured sensor response (resistance in Ω) for sensor j as the bacteria's headspace sample (combined with pure air carrier gas) passes over the sensors, R_{j0} is the baseline response for sensor j (at $t = 0$, measured in the presence of carrier gas only), and $r_{j,preproc}(t)$ is the scaled response retained for the following stages. No baseline manipulation was required on the FMS data.

3.2.3.2 Feature Extraction

Using all of the data in the resulting response curves would constitute an amount of information too large to use unaltered in a pattern recognition system. It is necessary to reduce the data set into a more compact form for subsequent processing. Ideally, this would be done in a manner that retains the information essential for representing and differentiating between the different bacteria species. Though a multitude of methods exist (some of which are more sophisticated than others [87][88]), the most common

approach (and the one adopted here) is to represent the time series for each sensor with a single value. For the MOS and CPS, the peak absolute value of $r_{j,preproc}(t)$ was retained as the feature of interest – for positively responding sensors, this value will be positive (representing a maximum) and for negatively responding sensors, this value will be negative (representing a minimum). On the FMS, the feature for each mass fragment “sensor” was obtained by finding the area under the intensity vs. time curve during the time interval for which the intensity was greater than or equal to half of its maximum value. Figure 3.5 illustrates the feature extraction process for all three instruments. Work was performed to investigate the value of other feature extraction methods (e.g., those based on the wavelet transform and a subsampling of the sensor response curves), but these were shown empirically to add little discrimination power to the overall system – for this reason, the conventional approaches described above were used.

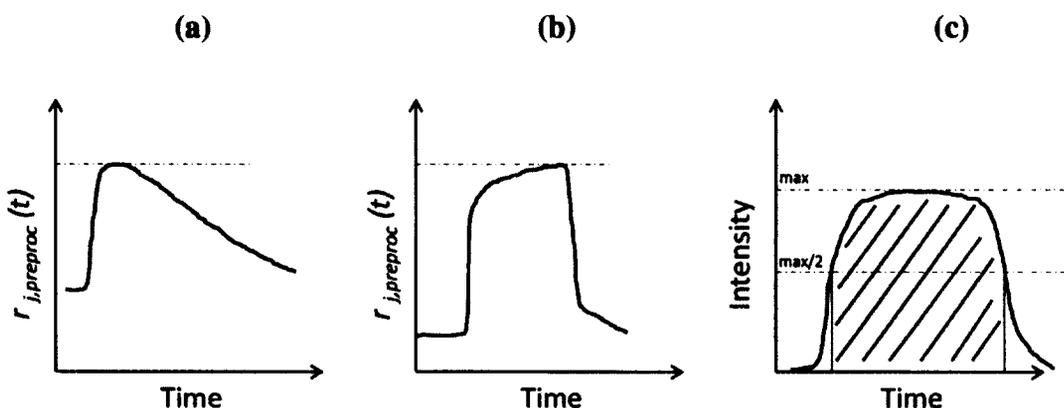


Figure 3.5: Illustration of the feature extraction process for the three e-nose instruments used in this thesis. (a) MOS based e-nose, (b) CPS based e-nose, (c) FMS based e-nose.

3.2.4 Dimensionality Reduction

Processing these feature vectors requires working in a high dimensional space and this is problematic by nature – the features are generally highly correlated (since the

sensors have overlapping sensitivities) and it becomes impossible to visualize the clusters. Furthermore, the number of training samples required to cover this space is prohibitively large (it must grow exponentially with the dimension of the space). Known as the *curse of dimensionality*, this is a significant hindrance in machine learning systems [89]. Almost invariably, e-nose systems perform a dimensionality reduction (DR) stage, wherein the number of feature vector components is decreased significantly from the original.

In this thesis, attention was restricted to three DR methods that form new features *from linear combinations of the original features*. Each method uses a different optimization criterion to form the linear coefficients used in the DR transformation matrix, as described below. Prior to this, however, an initial sensor selection was performed on the FMS e-nose. Upon visualizing the FMS raw data, there was only a small fraction of the 106 mass fragments that were responding with a discernibly strong intensity. The weaker values were eliminated by hard thresholding – those fragments for which none of the samples generated an intensity value greater than an empirically determined noise floor were eliminated, and the remaining fragments were kept. On the MOS and CPS, all sensors were retained.

3.2.4.1 Principal Components Analysis (PCA)

Ubiquitous in its use, PCA is an *unsupervised* DR method, meaning that the samples' category labels are not used. PCA represents the data set by creating a new basis from linear combinations of the original dimensions along which the scatter of the data points is greatest. The linear combination is defined by the PCA transformation matrix \mathbf{P} , which

defines the relationship between the original feature vector \mathbf{f} and its representation in the PCA space, \mathbf{f}_{PCA} according to $\mathbf{f}_{\text{PCA}} = \mathbf{P} \mathbf{f}$. The coefficients of \mathbf{f}_{PCA} are called the principal components and generally 2 or 3 are used to describe the data set [90].

3.2.4.2 Multiple Discriminant Analysis (MDA)

When discrimination between sample categories is desired, DR can be performed in a *supervised* manner. The category labels of the feature vectors are used to generate a linear transformation that reduces the scatter of samples within the same category and maximizes the separation of samples from different categories in the transformed space. MDA forms a projection (defined by the transformation matrix \mathbf{M}) from the original highly-dimensional feature vector space onto a much lower $(C - 1)$ - dimensional space based on an eigenvalue analysis (where C = number of categories), $\mathbf{f}_{\text{MDA}} = \mathbf{M} \mathbf{f}$. The resulting MDA components of \mathbf{f}_{MDA} are ordered in terms of their discrimination ability. MDA is a generalization of Fisher's linear discriminant analysis suitable for the multicategory (i.e. $C > 2$) case [90].

3.2.4.3 Uncorrelated Linear Discriminant Analysis (ULDA)

The criterion function in MDA requires a *non-singular* scatter matrix. ULDA (also a supervised DR method) is useful in instances where the scatter matrices of the samples are singular (and thus non-invertible) – this can arise when there are many more features than samples (called *undersampling*), or when the features are highly correlated. Its formulation uses a criterion function based on the pseudo-inverse of the scatter matrices.

When the ULDA transformation matrix U is applied to the original features, $\mathbf{f}_{\text{ULDA}} = U \mathbf{f}$, the new features in the transformed space, given by \mathbf{f}_{ULDA} , are uncorrelated [91].

3.2.5 Classification and Performance Characterization

To quantify the discrimination performance of the e-nose in experiments reported throughout the thesis, a classifier is used to assign a category label to a new sample (represented by its feature vector) using a previously trained model. The e-nose literature entails a wide variety of classifier methods, among these include k -nearest neighbours, MLPs, radial basis functions, support vector machines, and decision trees [86][84]. An important factor for any classifier is to reduce the potential for *overtraining*. This occurs when the classifier does not generalize well for novel (i.e. not yet encountered) test samples, despite performing very well on the training samples. In this thesis, the linear discriminant analysis (LDA) [92] classifier was used, also common in the e-nose literature [86]. This classifier is a simple one, generating linear decision boundaries. This decision was deliberate – the primary interest was to assess the discrimination performance of the sampling method based on the quality of the feature vectors (and not by training an overly complex classifier, such as an MLP which has the potential to be overtrained [93]).

In order to validate the classifiers (and to estimate performance), this thesis employs the “leave-one-out” (LOO) cross validation technique [84]. With LOO, a single iteration uses all but one of the samples ($N - 1$, the training set) to train the classifier (i.e. determine the decision boundaries). The remaining sample (the test set) is withheld and presented for classification. The output category is computed and compared to the true

category, yielding a classification accuracy (CA_i) of 100% or 0% for that iteration i . This process is iterated over all N possible partitions (so that each sample is used as a test point once), and the overall classification accuracy CA is averaged.

Throughout the thesis, there are instances where classification was performed on feature vectors that had their dimensionality reduced (with either of the methods described in Section 3.2.4) prior to classification. In these cases, the DR was performed using all of the samples (prior to splitting samples between training and testing sets), an important step that helps to avoid overtraining that would result in overly optimistic CA values.

3.3 Experiment #1 – Exploratory Analysis of Dense Cultures Using Three E-nose Modalities

3.3.1 Experiment #1: Bacteria Sample Preparation and Sampling Protocol

As was already stated, the intent of Experiment #1 was to characterize the ability of the three e-nose instruments (Section 3.2.2) to discriminate dense cultures and select those appropriate for further work. The bacteria preparation described in Section 3.2.2 resulted in dense cultures of 1.4×10^9 cells/mL for *E. coli* and 6.0×10^8 cells/mL for *Listeria* (the suspension medium was LB broth). For each e-nose, three samples from each category – *E. coli* in LB broth (label: E), *Listeria* in LB broth (label: L), and LB broth control (label: B) – were processed. For MOS and FMS e-noses, 1mL of bacterial suspension was transferred by pipette into 10mL vials, and vials were immediately sealed with polytetrafluoroethylene (PTFE)-lined caps to prevent contamination. For the CPS e-nose (which did not support sampling from sealed vial), sniffs were taken by positioning

the sampling snout roughly 5cm above the surface in a flask containing 100mL of dense culture. In all cases, the replicates were sampled in alternating order (E, L, B, E, L, B ...) and for each instrument, all vials were processed on the same day.

3.3.2 Experiment #1: Electronic Nose Processing

For the MOS e-nose, the vials were heated to 50°C for 300s to generate headspace. Immediately following this, 1mL of headspace (at a rate of 1mL/s) was injected into the MOS sensor chamber. For the FMS e-nose, the headspace generation time/temperature was 900s/60°C and 4mL of headspace (at a rate of 0.1mL/s) was injected into the quadrupole mass spectrometer. The CPS e-nose did not support headspace generation.

3.3.3 Experiment #1: Results

Figure 3.6 demonstrates the ability of the MOS, FMS and CPS e-noses to discriminate between the dense bacteria cultures. Each point in these plots represents a single bacteria sample in the dimensionality-reduced PCA feature space.

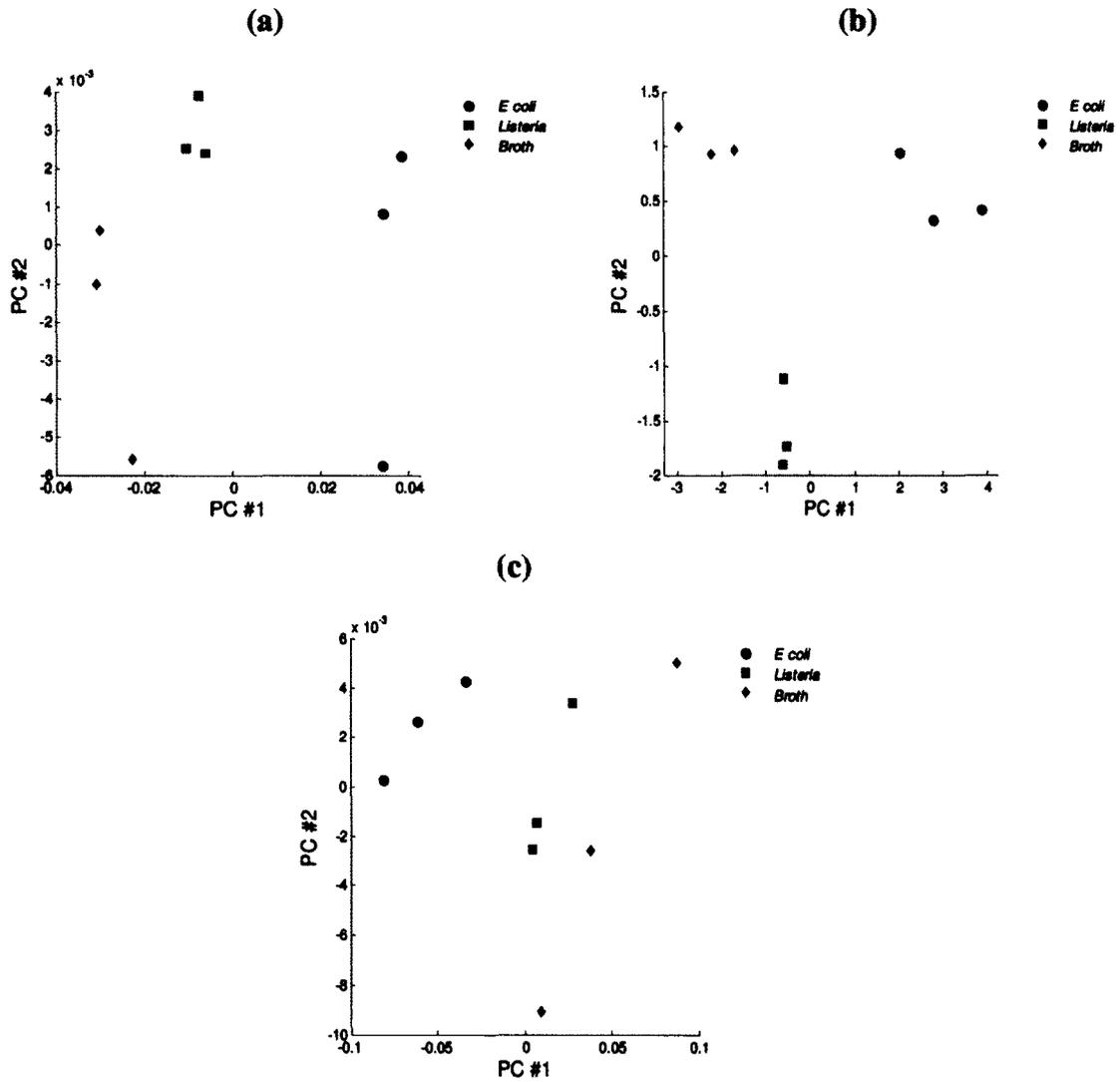


Figure 3.6: Experiment #1: PCA plots illustrating the ability of (a) MOS, (b) FMS and (c) CPS e-noses to discriminate between the dense bacteria cultures and nutrient broth. Each point represents a bacteria sample in the dimensionality-reduced PCA feature space.

3.3.4 Experiment #1: Conclusions and Discussion

The separation between E, L, and B clusters (relative to the size of the cluster) is more pronounced for the MOS/FMS than it is for the CPS system. It should be noted that the HS100 autosampler used by the MOS and FMS instruments allows for precise settings of both headspace generation and injections, but such control was not easily achieved with the CPS instrument. Further refinements in testing with the CPS would likely lead to better performance. Undoubtedly, variations introduced by manual sampling accounted, to some degree, for the inferior results achieved with this instrument. Initial testing with the CPS system was attempted using a sample size of 1 mL in a 10 mL (as was used with the MOS/FMS instruments). Without the same combination of controlled sampling, agitation, and heating, the headspace generated was not strong enough to give a reliable result. For these reasons, only MOS and FMS e-noses were used for subsequent experiments (which involved reducing the bacteria concentration significantly).

3.4 Experiment #2 – Discrimination of Bacteria with Varying Concentrations (Dilutions Done Prior to Sampling)

3.4.1 Experiment #2: Bacteria Sample Preparation and Sampling Protocol

This experiment investigated the ability of MOS- and FMS-based instruments to discriminate bacteria from less dense cultures (as would be available earlier in the culturing process – see Figure 1.1). Varying concentration was achieved by serial dilution with additional LB broth to achieve levels of 10^8 , 10^7 and 10^6 cells/mL. Bacteria samples were processed on three separate days – one for each concentration. All bacteria

were stored at 4°C prior to sampling to restrict further growth. There were 18 samples processed per day – six *E. coli*, six *Listeria*, and six LB broth. In this experiment, each dilution forming a new concentration was performed immediately prior to running the samples at that concentration. A total of 54 samples were processed.

3.4.2 Experiment #2: Electronic Nose Processing

In order to elicit a stronger response (as compared to Experiment #1) for these less dense samples, a couple of changes were made: a) the sample aliquot was increased to 2mL (for both MOS and FMS), and b) the headspace generation temperature/time was increased to 100°C/900s (for both MOS and FMS). The headspace injection volume and rate remained the same as was used in Experiment #1 (Section 3.3.2).

3.4.3 Experiment #2: Results

After performing serial dilutions of the dense cultures to create different concentrations, Figures 3.7 and 3.8 demonstrate the effect on the clustering results (this time showing both unsupervised PCA and supervised ULDA) for the MOS and FMS e-noses, respectively. The MOS results (Figure 3.7) show that a) the clusters are more distinct at higher concentrations (i.e. there is a larger separation between groups and less intra-cluster scatter) and b) the supervised DR methods give superior results. Discrimination between bacteria categories is possible using the MOS instrument and supervised techniques, down to a concentration of 10^6 cells/mL. On the FMS, when the noise floor was adjusted to use the same number of retained features as the MOS system (12), the results were generally in agreement (Figure 3.8). However, the ULDA

representation generated for the FMS does not show the same degradation in cluster width as a function of decreasing concentration (as is evident with the MOS results).

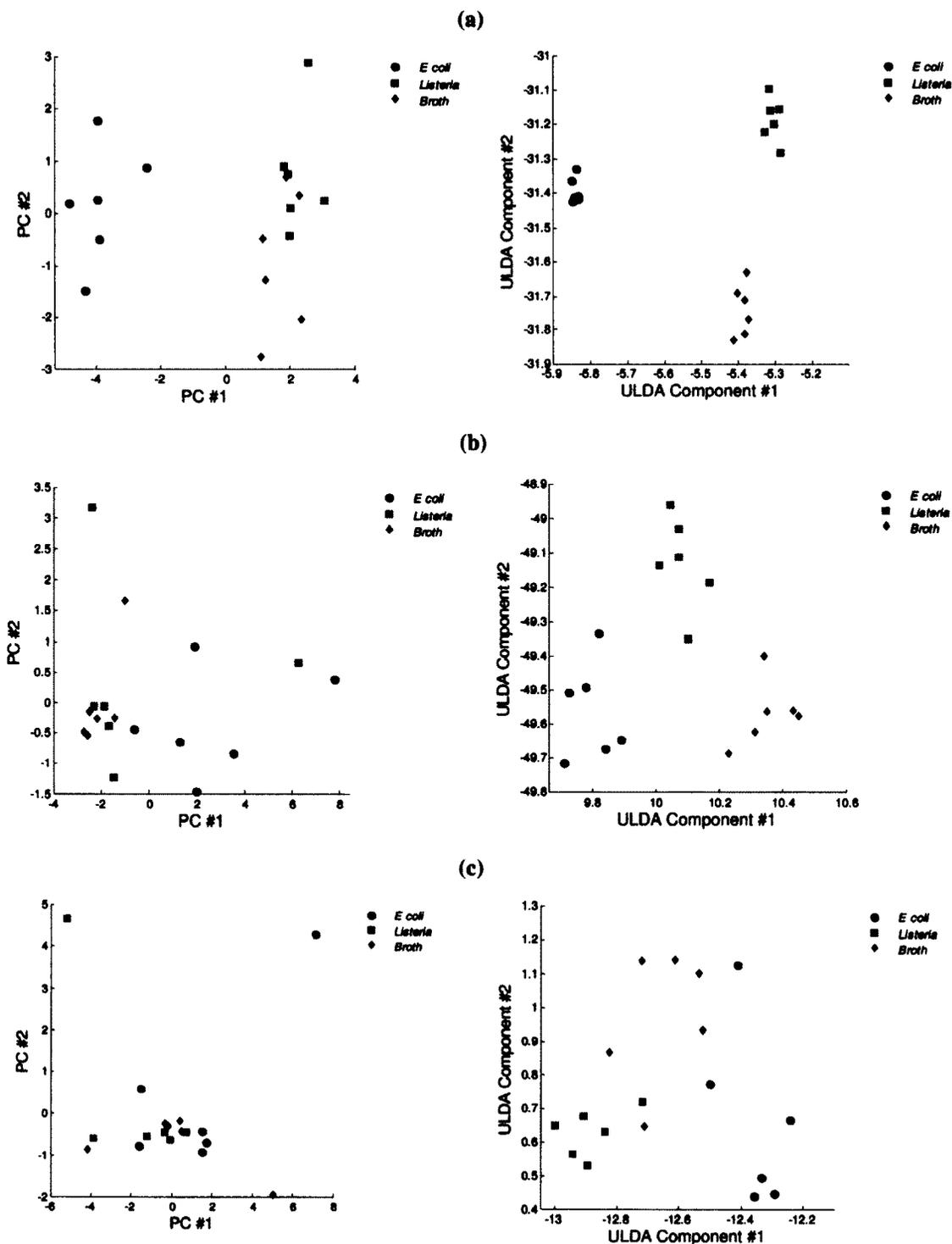


Figure 3.7: Experiment #2: Clustering results illustrating the ability of MOS-based e-nose to separate bacteria samples at concentrations of (a) 10^8 , (b) 10^7 , and (c) 10^6 cells/mL using dimensionality reduction (LEFT: PCA, RIGHT: ULDA).

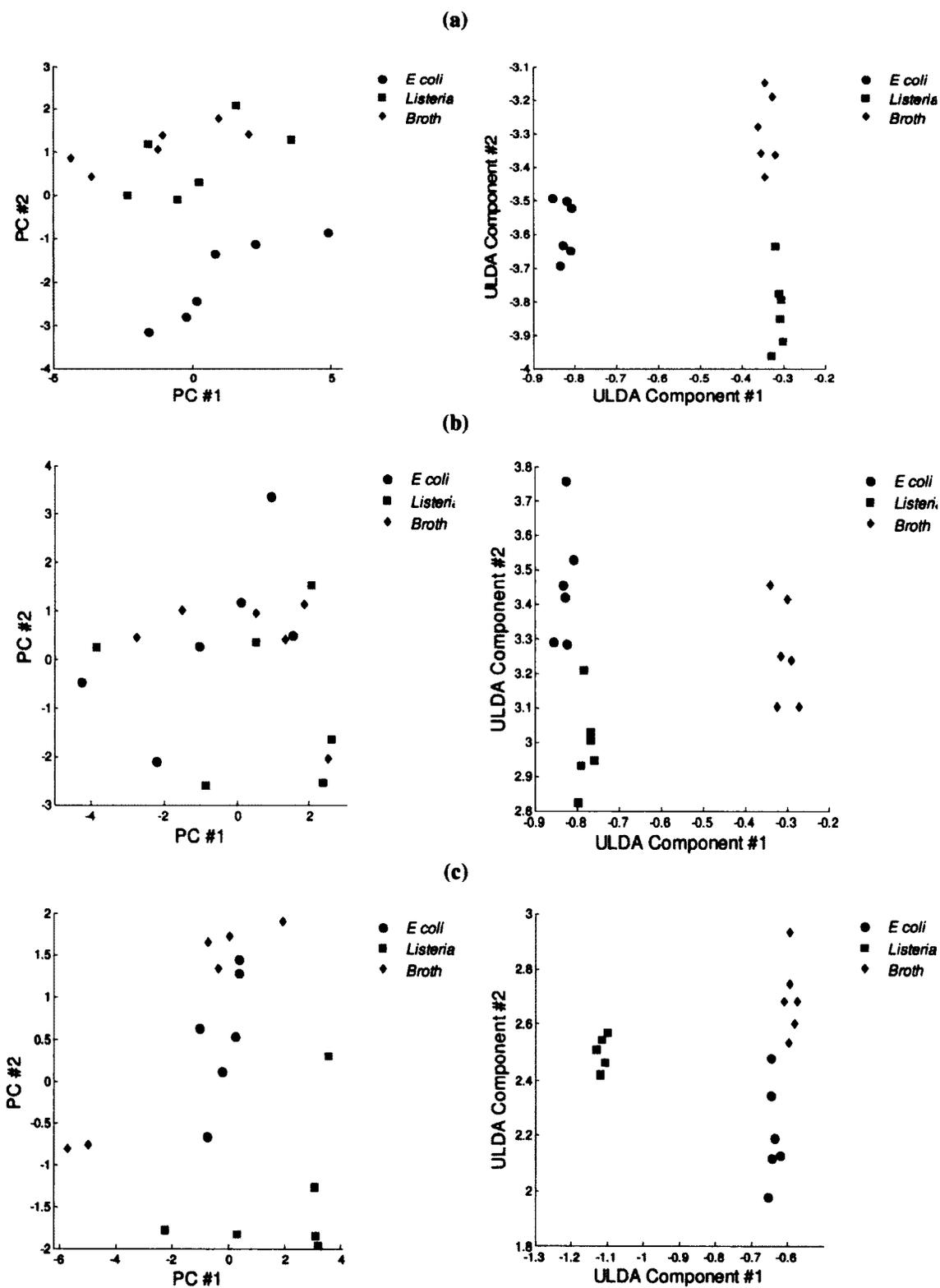


Figure 3.8: Experiment #2: Clustering results illustrating the ability of FMS-based e-nose to separate bacteria samples at concentrations of (a) 10^8 , (b) 10^7 , and (c) 10^6 cells/mL using dimensionality reduction (LEFT: PCA, RIGHT: ULDA).

3.4.4 Experiment #2: Conclusions and Discussion

The results above are intuitive, since suspensions at the higher concentrations would be expected to produce a VOC pattern specific to the volatile metabolic products of each bacteria type. At the lower concentrations, there are fewer bacteria present in the sample, so the odour pattern more closely resembles that of the underlying medium (LB broth). However, these results also illustrate the performance when *only the samples for a given day* (or equivalently in this protocol, a single concentration) are shown. When all of the samples (encompassing all concentrations) are pooled together, grouped in the same manner (E, L, B) and processed as above, some interesting properties arise. Figure 3.9(a) gives the results for the MOS instrument. This plot shows a high degree of overlap among samples in the different groups, thus no obvious clustering is apparent (based on bacteria type). Figure 3.9(b) plots the same points, but in this case they are individually labeled based on their bacteria type and concentration. The property that best separates the samples in this case is not the *bacteria type*, but rather the *concentration* of the samples (or equivalently, the time at which they were processed since for a fixed concentration, all samples occur in the same run). In this analysis, only PCA (an unsupervised technique) was used (to avoid forcing any specific groupings through supervision). Similar results were seen on the FMS. In order to better understand this phenomenon, the effect of varying bacteria concentrations was eliminated by considering *only the LB broth samples* from each of the separate runs. Figure 3.10 shows the PCA results when these samples are analyzed (using both FMS and MOS). In these plots, the LB broth samples are labeled based on the run number (or equivalently, the day on which they were processed). In the absence of bacteria, the broth samples measured on different

days cluster in different areas of the PCA space. This demonstrates that the volatiles associated with the nutrient broth varied during the entire measurement period (thereby confirming the drift effect described earlier).

To summarize, the results above suggest that using this sampling protocol, separation between bacteria classes at a fixed concentration level is possible (using supervised DR methods), down to 10^6 cells/mL. When samples of all concentrations (representing different processing times) are grouped together, however, the ability to do so is lost.

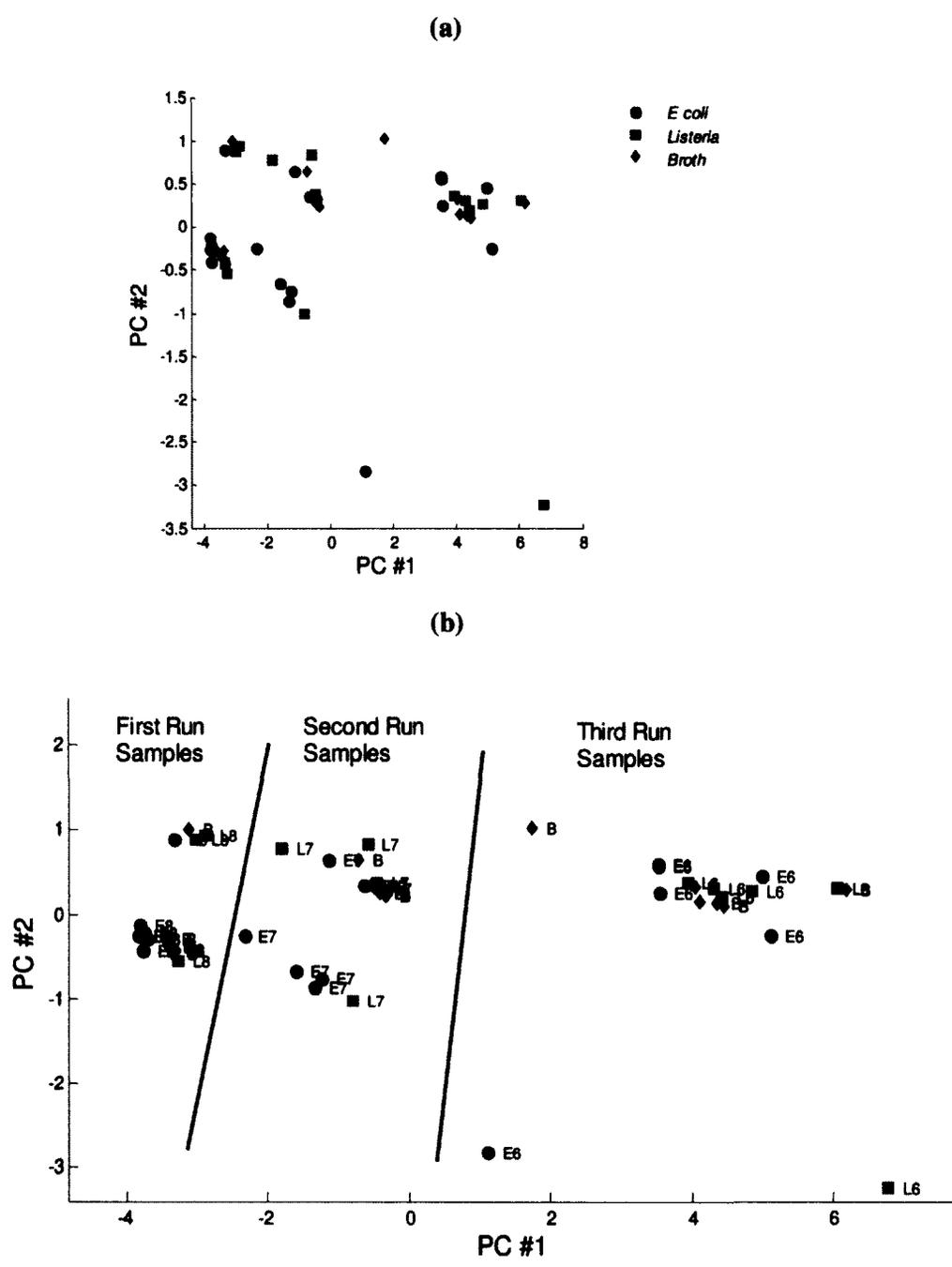


Figure 3.9: Experiment #2: (a) PCA clustering for MOS-based e-nose when all three concentrations (representing different processing days) are pooled. This plot shows a high degree of overlap among bacteria and broth samples. (b) PCA clustering of the same points, individually labeled based on their bacteria type and concentration (e.g., E7 is *E. coli* at 10^7 cells/mL). It is evident in (b) that there is an inherent separation based on the day in which the samples were processed.

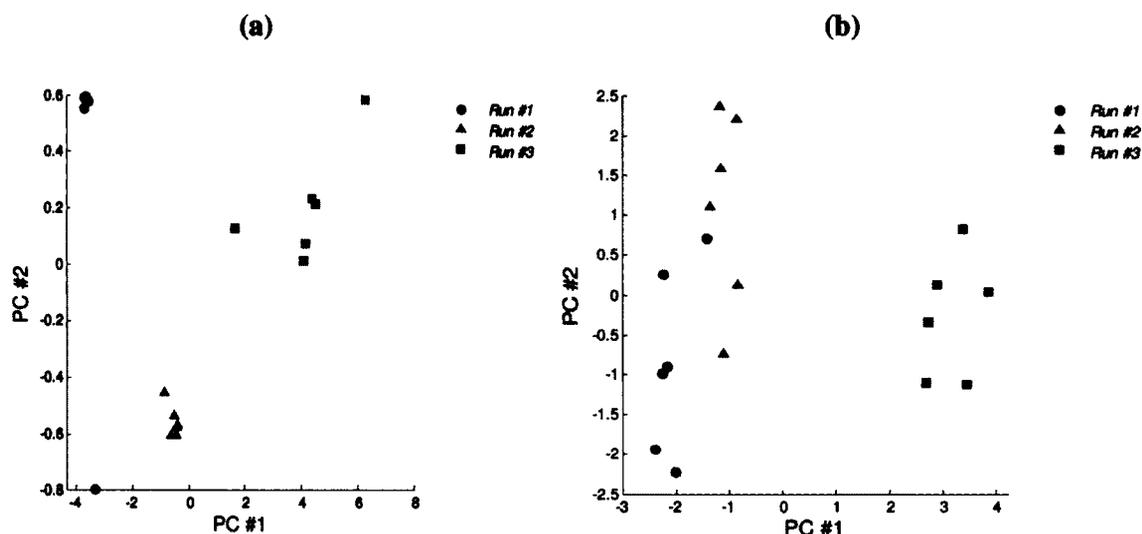


Figure 3.10: Experiment #2: PCA clustering for LB broth samples only, with groupings based on the processing day: (a) MOS-based e-nose, (b) FMS-based e-nose.

3.5 Experiment #3 – Discrimination of Bacteria with Varying Concentrations (Dilutions at $t=0$)

3.5.1 Experiment #3: Bacteria Sample Preparation and Sampling Protocol

This experiment was performed in an attempt to address the time-varying effect seen in Experiment #2 (which made it impossible to assess the e-nose’s ability to discriminate based on species/concentration). In Experiment #3, all serial dilutions were performed *at the beginning of the experiment*. Instead of only including a single concentration on each processing day (Experiment #2), samples from all species/concentrations were included in a single run, and processing was done over five separate days during a week-long period. There were 30 vials included in each run, consisting of 10 LB broth, 10 *E. coli* (including 2 vials each at five concentration levels of 10^7 , 10^6 , 10^5 , 10^4 and 10^3 cells/mL) and 10 *Listeria* (again, including 2 vials each at five concentration levels of 10^7 , 10^6 , 10^5 , 10^4 and 10^3 cells/mL). A total of 150 samples were processed. Spreading out samples of

varying concentrations represents a more realistic testing scenario (since the identification of bacteria may need to be performed at any arbitrary time). Using an increased number of samples and lower concentrations will provide a better idea of the detection threshold of the system.

3.5.2 Experiment #3: Electronic Nose Processing

All headspace generation and injection parameters used were identical to those used in Experiment #2 (Section 3.4.2).

3.5.3 Experiment #3: Results

The alternate grouping used in this experiment allowed an investigation of the effect of including samples from each bacteria/concentration combination for a single processing run (spanning several hours only, when the LB broth drift is presumed to be negligible). The LB broth drift clearly prevents the ability to detect and discriminate the bacteria classes (when all samples are considered), as shown in Figure 3.11. It is evident from this figure that it is only *Listeria* at the highest concentrations (10^7 and 10^6 cells/mL) that can be separated (using the supervised methods).

Because bacteria discrimination from more than one run presents problems, the data was limited to that of a single run only (Run #3 chosen arbitrarily). This particular run did not include MOS processing, so it was quite fast (roughly 3 hours), and it is assumed that the LB broth drift effect was not significant during this time. Figure 3.12 shows the results obtained using ULDA for two possible data groupings – one based on bacteria type alone, and the other based on bacteria type *and* concentration. Overall, the

separation based on bacteria type is superior to that of Figure 3.11 (which includes data from all runs).

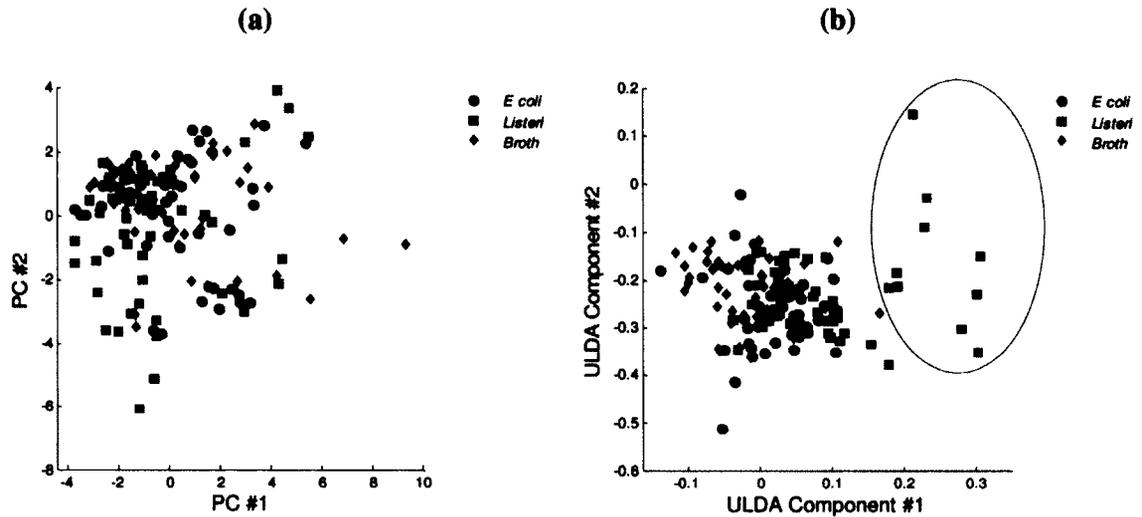


Figure 3.11: Experiment #3: Clustering results illustrating the ability of FMS-based e-nose to separate bacteria and broth samples at concentrations varying from 10^7 cells/mL down to 10^3 cells/mL: (a) PCA, (b) ULDA. In (b), the circled samples correspond to *Listeria* samples at the highest concentrations (10^7 and 10^6 cells/mL).

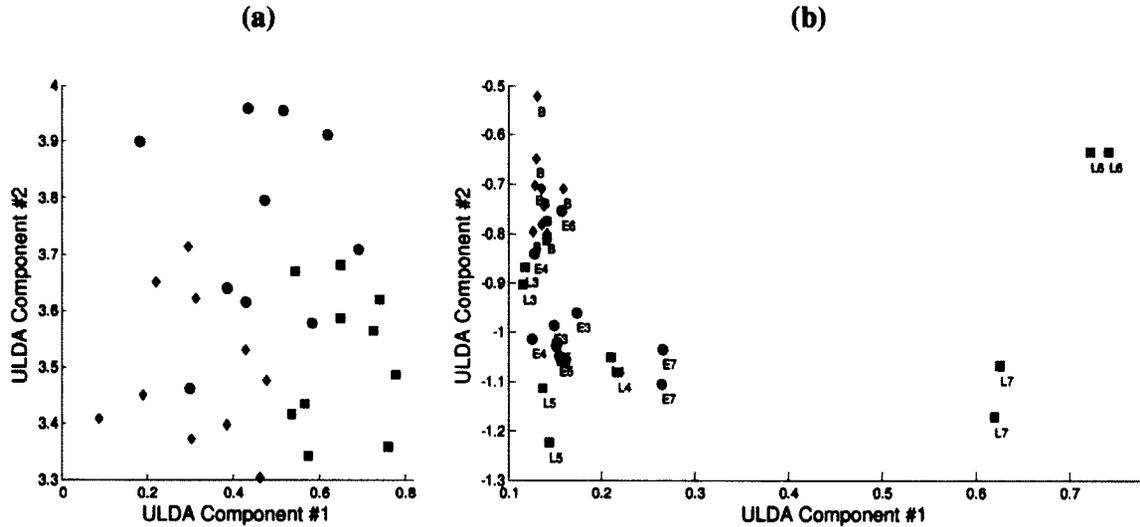


Figure 3.12: Experiment #3: ULDA clustering results for two category groupings (Run #3 only): (a) based on bacteria type alone, and (b) bacteria type and concentration (e.g., E5 is *E coli* at 10^5 cells/mL, L3 is *Listeria* at 10^3 cells/mL).

3.5.4 Experiment #3: Conclusions and Discussion

When the category grouping is done over all combinations of bacteria type and concentration (11 categories total), the results in Figure 3.12(b) indicate that in the absence of significant LB broth drift, there is a *suggestion* of discernability between the individual classes. However, one cannot conclude this because of the limited data available. Since each run can hold a maximum of 30 samples (10 of which are LB broth), the 10 unique bacteria/concentration categories are represented by only 2 data points each. Nonetheless, this result motivates further study into whether, in the absence of LB broth, thresholds for detection and discrimination can be significantly lowered.

The drift problem is well acknowledged in the e-nose literature, and there have been a wide variety of drift compensation techniques employed to deal with it [84][86]. However, these techniques generally assume that the drift is due to slow variation in the sensor baselines or background media (in this case, the broth). The results presented previously (Figure 3.10) illustrate that this is not the case.

3.6 Overall Summary and Conclusions

In this chapter, an exploratory analysis demonstrated that discrimination of dense bacterial cultures ($\sim 10^9$ cells/mL) in LB broth was possible with three e-nose instruments (Experiment #1). The most sensitive of these, MOS and FMS, were then used as this concentration was dropped down to 10^6 cells/mL. Evidence of discrimination was possible on a single day, but combining samples over several days uncovered a time-varying drift effect that hindered overall classification performance (Experiment #2). It is

strongly suspected that this is due to changing properties of the LB broth volatiles over time. Because LB broth was used as both nutrient medium and dilution agent, it forms the dominant component of the samples, especially at lower concentrations. Experiment #2 demonstrated that variations in the features that are due to bacteria type (while evident during a single run) are overwhelmed by larger variations in the time-varying properties of the LB broth. Processing still lower concentrations (down to 10^3 cells/mL) on the same day suggests that discrimination between *Listeria* and *E. coli* is possible in these ranges (Experiment #3), but further validation needs to be performed to confirm this.

The bacteria samples used in this chapter would arise relatively late in the overall foodborne pathogen identification process, after colony purification has occurred. Figure 3.13 highlights the stage at which this sample type is available. As such, the reason to employ e-nose analysis at this stage would not be to decrease overall inspection time. Despite this, it could be used as an alternative technique (requiring little sample preparation and very low cost), to conjecture as to the nature of the contamination bacteria, which could then be confirmed with conventional methods.

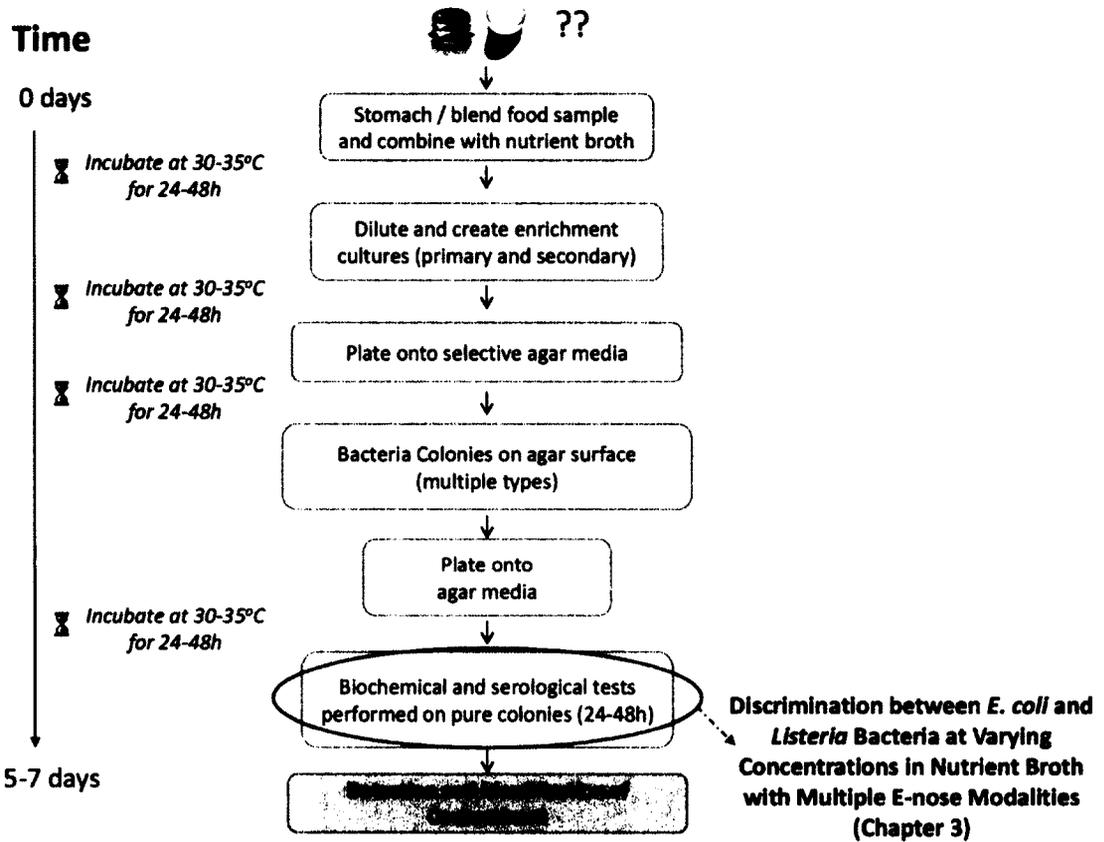


Figure 3.13: Simplified view of conventional methods. Highlighted block indicates the stage at which the sample type considered in Chapter 3 is available.

Chapter Four

Discrimination between Single Colonies of *E. coli*, *Listeria* and *Enterococcus* Bacteria with a Metal Oxide Sensor-based Electronic Nose

In Chapter 3, one of the shortcomings of using nutrient broth as the bacteria medium in an e-nose measurement system was illustrated – the presence of drift. In this chapter, a sampling and measurement system is described that is designed to circumvent this problem by changing the sample type entirely; that is, removing the nutrient broth by instead using individual bacteria colonies. This has two advantages: a) the individual colonies are available earlier in the food inspection process (when compared to the dense culture broth suspensions used in the previous chapter) and b) it has the potential to eliminate the confounding effects of drift due to nutrient broth.

The work described in this chapter has been disseminated in [22].

4.1 Background

In Section 2.4, literature was presented that reported the success of using various e-nose sensors (in concert with pattern classification algorithms) to identify bacteria types. Sample types and preparation varied amongst studies, but the following types of samples were found to predominate:

1. real food samples that had been spiked (inoculated) with a single contaminating bacterial species (e.g., [72]),
2. numerous colonies of the same bacteria grown on agar in a Petri dish (e.g., in [94]), or

3. dense bacteria cultures in a growth medium, where the originating bacteria came from either a food sample or a lab stock source (similar to that described in the previous chapter).

These samples are useful to demonstrate the potential of e-nose based bacteria discrimination, but there are shortcomings with using each of these approaches in food safety. In the first case, spiked food samples do not realistically represent what might arise in the field, since real food samples can be tainted with *multiple* types of bacteria in very low doses (of course, this information is unknown *a priori*). The food inspection process deals with this as follows (see Figure 4.1). Once the food has been homogenized or blended (and following a suitable primary enrichment incubation period as shown in Figure 1.2), colonies of different bacteria may emerge on the plate. In order to identify each of these bacteria types, an individual colony will be removed from the plate, re-enriched (usually with a selective nutrient broth), re-plated and then left for an additional 24-48h incubation phase. After this stage, multiple colonies of the same bacteria type would emerge on the plate – it is at this point that the second sample category above would be available. It is only after this stage that each species has been suitably isolated and purified to make dense cultures of that bacterial type possible (the sample type represented in the third category above).

These observations, coupled with the problem of drift mentioned in the chapter preamble, motivated an attempt at e-nose analysis with a novel sample type – *single bacteria colonies taken directly from the agar surface*. The hypothesis was that since the amount of nutrient residue in this sample would be minimal, the drift problem should be reduced. Additionally, the proposed sampling method provided the potential to identify

bacteria on a per-colony basis, so the (unrealistic) assumption that the food is contaminated by a single bacteria type would be avoided. Furthermore, it has the potential advantage of avoiding the significant delays inherent to the additional culturing/plating stage required to purify the colonies with the conventional methods (see Figure 4.1).

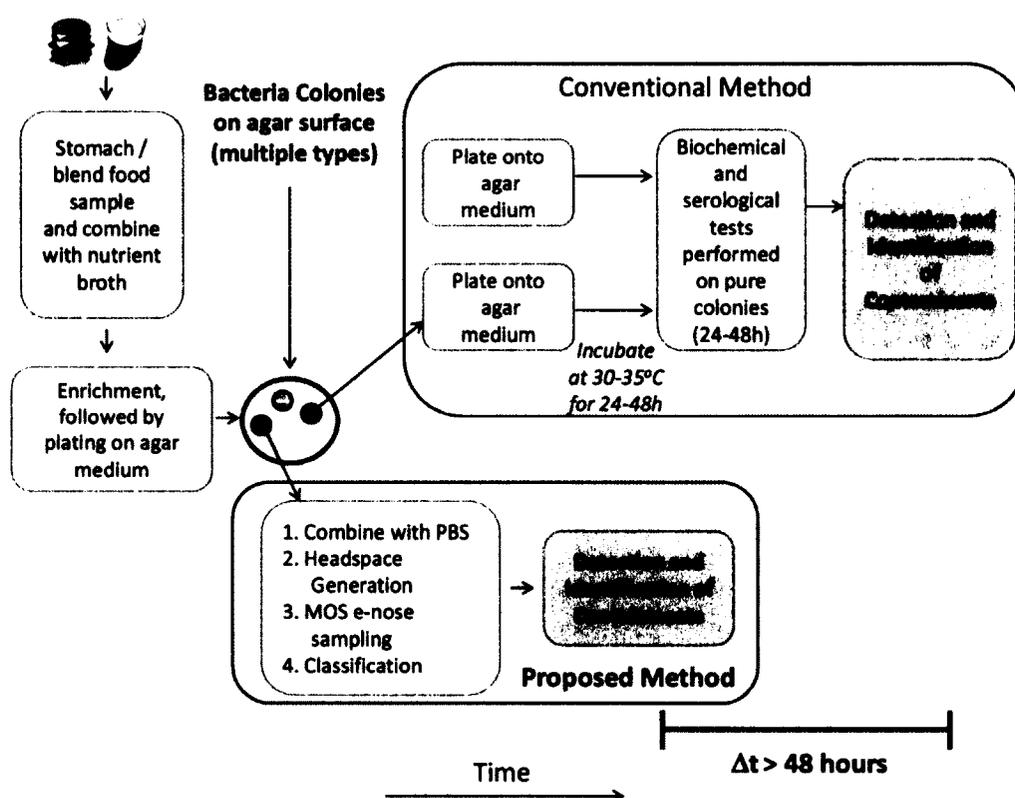


Figure 4.1: Illustration of the manner in which multiple colonies can grow on a plate during food testing. Conventional methods must obtain pure colonies and further incubate each type prior to biochemical and serological testing. The e-nose method described in this chapter has the potential to eliminate the need for the latter stages.

In this chapter, two experiments are described. Experiment #1 represents an initial assessment of the ability of the MOS-based e-nose system to discriminate between single colonies of two different bacteria types (*E. coli* and *Listeria*) grown on agar. Experiment #2 provides a validation of the methods demonstrated in Experiment #1 in several

respects: a) a larger selection of bacteria is used (four instead of two), b) the growth medium used in agar preparation is controlled, and c) the size of the colonies is controlled. In addition, Experiment #2 investigates whether an increase in discrimination performance can be achieved by acquiring multiple consecutive e-nose responses (or *sniffs*) of the same sample, as the literature has suggested [95].

This work in this chapter is organized by first describing the methods used that are common to *both* experiments (Section 4.2). Following this, method details unique to Experiment #1 (mostly relating to bacteria sample preparation) are described, and the results and discussion of Experiment #1 are presented (Section 4.3). These results provide the motivation for the different sampling and signal processing methods employed in Experiment #2, then followed by the results and discussion of Experiment #2 (Section 4.4). Section 4.5 presents an overall summary and conclusions of this chapter.

4.2 Methods

The methods described below are common to *both* experiments – these include e-nose sampling parameters, headspace generation parameters, data processing and classification.

4.2.1 Electronic Nose Analysis

The experiments described in this chapter used the AlphaMOS FOX 3000, the MOS-based e-nose whose principle of operation has been described earlier (Chapter 3). The HS100 headspace generation parameters and e-nose instrument parameters that were used are shown in Table 4.1.

	Parameter Value
Headspace Generation Parameters	
Agitation Time	300s
Agitation Temperature	50°C
E-nose Sampling Parameters	
Injection Volume	5mL
Syringe Temperature	55°C
Injection Speed	1mL/s
Carrier Gas Flow Rate	150mL/min
Delay Between Samples	1200s

Table 4.1: MOS e-nose parameters used in single colony bacteria discrimination experiments.

4.2.2 Data Processing - Baseline manipulation

The raw sensor resistance responses $\{R_j(t), j = 1, \dots, 12\}$ were measured, then preprocessed using fractional baseline manipulation forming $r_{j,preproc}(t)$ (Equation 3.1) to eliminate the effect of short term drift in sensor baseline.

4.2.3 Data Processing - Feature Extraction and Dimensionality Reduction

It was necessary to extract meaningful features from the $\{r_{j,preproc}(t), j = 1, \dots, 12\}$ curves for subsequent processing. Ideally, these features would retain only the essential information that is required to represent and differentiate the samples. With this e-nose sensor type, it is typical to capture the peak absolute value response for each sensor, $f_{j,max}$ (see Section 3.2). For positively responding sensors, this value will be positive (representing a maximum) and for negatively responding sensors, this value will be negative (representing a minimum). This represents a single value that provides only steady-state information. This feature alone does not provide any information as to the *dynamics* of the response and there is recognition in the e-nose literature that incorporating this information can be useful [88][96]. For this reason, an additional

feature from each of the $\{r_{j,preproc}(t)\}$ curves were obtained by averaging the first difference of the sampled $r_{j,preproc}(t)$ curves in the transient region, $5 < t < 20s$ (which at a sampling interval $T_s = 0.5s$ equates to indices $k = 10, 11, \dots, 40$) to form an additional feature called $f_{j,diff}$, forming an approximation to the slope:

$$f_{j,diff} = \frac{1}{31} \sum_{k=10}^{40} (r_{j,preproc}((k+1)T_s) - r_{j,preproc}(kT_s)) \quad (4.1)$$

The feature extraction process is shown in Figure 4.2. To summarize, the i^{th} bacteria sample is represented with a feature vector which has 24 components (two for each sensor):

$$\mathbf{f}_i = [f_{1,max} \ f_{1,diff} \ f_{2,max} \ f_{2,diff} \ \dots \ f_{12,max} \ f_{12,diff}] \quad (4.2)$$

For the same reasons outlined in Section 3.2.4, both unsupervised DR (using PCA) and supervised DR (using ULDA) were employed.

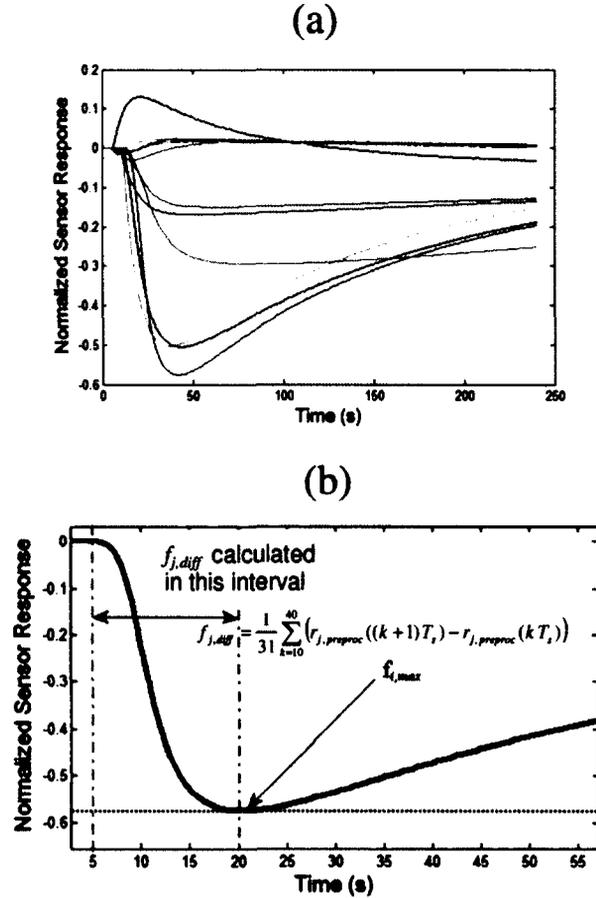


Figure 4.2: Plot illustrating: (a) the 12 sensor responses for an arbitrary sample (the normalized response is shown) and b) example of extracted features for an arbitrary sensor. Two features are extracted per sensor: the maximum absolute value (labelled $f_{j,max}$) and a derivative estimate (labelled $f_{j,diff}$). For the negatively responding sensors shown in (b), both $f_{j,max}$ and $f_{j,diff}$ will be negative values.

4.2.4 Classification and Performance Characterization

To quantify the performance of this novel sampling method in discriminating the *E. coli* and *Listeria* samples, a linear discriminant analysis (LDA) classifier was used and LOO cross validation was used to estimate classification accuracy (see Section 3.2.5).

4.2.5 Confidence in Classification Results

The purpose of the classification stage is to assign a category label to a new measurement (based on that measurement's feature vector and the classifier model

derived from the training samples). Certain types of measurements (e.g., outliers, or sample types not encountered during training) may generate feature vectors that are significantly different from those present in the training set. While classifiers are able to predict the category label of such a sample, it is sometimes more suitable to use the “reject” option – meaning that instead of assigning a target category of questionable validity, the sample is deemed to be *unclassifiable*. This is particularly true in biomedical applications, where it is appropriate for classifier predictions to be made only when they are sufficiently reliable [97]. Using a classifier with a reject option requires a numerical measure of the confidence in a classification result. In this thesis, two such measures were considered:

1. **Posterior probabilities** – A very intuitive estimate of confidence is available for classifiers that, for a given measurement represented with feature vector \mathbf{f}_i , provide a estimate of the posterior probabilities for each class c , $p(\omega_c | \mathbf{f}_i)$, $c = 1, 2, \dots, C$ (where of course the sum of the probabilities add up to 1):

$$p(\omega_1 | \mathbf{f}_i) + p(\omega_2 | \mathbf{f}_i) + \dots + p(\omega_C | \mathbf{f}_i) = 1 \quad (4.3)$$

LDA (used in this thesis) is one such classifier. High confidence would be associated with those input patterns that generate $p(\omega_c | \mathbf{f}_i)$ values that are significantly higher for a single category than for all others. This can be achieved by choosing a threshold, T_p . If any of the C posterior probability estimates fall below this threshold, \mathbf{f}_i is rejected as unclassifiable [98][99]. This can be expressed as:

$$\text{Reject } \mathbf{f}_i \text{ if } p(\omega_c | \mathbf{f}_i) < T_p \quad \forall c = 1, \dots, C \quad (4.4)$$

2. Silhouette Coefficients – An alternative to the above can be developed by using the clustering behaviour of the input patterns in the feature space (either original or dimension-reduced). If a new input sample is found to lie in the middle of a cluster that defined a particular category k during training (while sufficiently far removed from other clusters), that sample should be assigned the k^{th} category label with a high degree of confidence. Otherwise, the input pattern lies closer to the boundaries between categories (which would raise ambiguity) and the confidence measure associated with that measurement should be penalized as a result. The use of *silhouette coefficients* (introduced by Rousseeuw [100]) can be used to accomplish this. Briefly, the following quantities are calculated for sample i (represented by \mathbf{f}_i) that has been assigned by a classifier to category k :

- The average distance between \mathbf{f}_i and all other feature vectors that define category k . This measure is called $a(i)$.
- For all other categories $c = 1, \dots, C$ (where $c \neq k$), the average distance between \mathbf{f}_i and the feature vectors belonging to the other categories. These $(C - 1)$ measures are called $d(i,c)$, $c \neq k$ and from this set, the minimum among the $d(i,c)$ values is kept. This value, called $b(i)$, represents the distance from \mathbf{f}_i to the closest neighbouring cluster.

The calculation of these quantities is illustrated in Figure 4.3. In this thesis, Euclidean distances were used. From these, the silhouette coefficient $s(i)$ is calculated as follows [100]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad -1 \leq s(i) \leq 1 \quad (4.5)$$

The following cases are of interest:

- Samples lying very close to the center of a tight cluster (and distant from the others). For these, $a(i)$ will be small relative to $b(i)$, and the $s(i)$ value will be correspondingly high (the maximum value is 1). These samples should have a *higher* degree of confidence associated with the classifier result.
- Samples closer to the boundaries between clusters. These will have increased $a(i)$ values, lower $b(i)$ values and correspondingly lower $s(i)$ values (the minimum value is -1). These samples should have a *lower* degree of confidence associated with the classifier result.

To conclude, high confidence should be associated with those input patterns that generate higher $s(i)$ values. By using a threshold T_s , it is possible to reject any samples falling below this threshold as unclassifiable, which can be expressed as:

$$\text{Reject } f_i \text{ if } s(i) < T_s$$

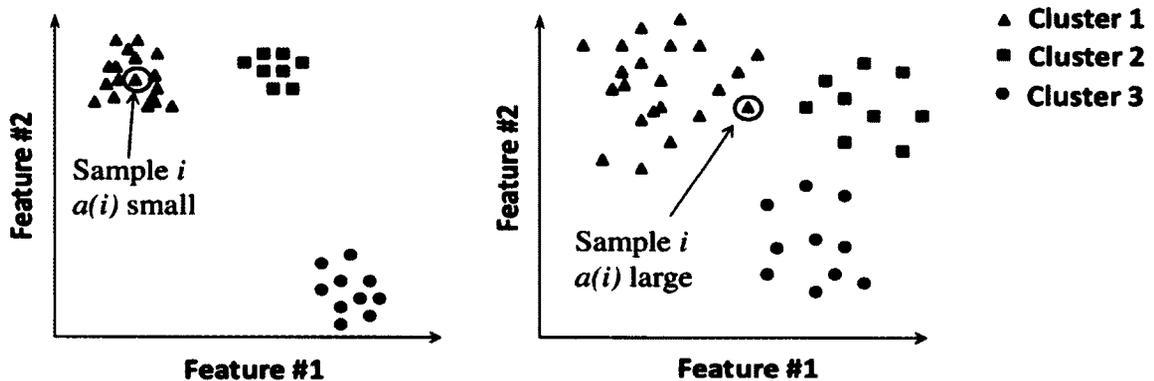


Figure 4.3: Demonstration of the calculation of the silhouette coefficient for a sample i (belonging to Cluster 1). Samples in the same category are used to define $a(i)$, while other categories are used to define the $d(i,c)$ values. The minimum of these is selected as $b(i)$. In the clustering on the left, $b(i)$ will be calculated for Cluster 2 (the closest), and since $b(i)$ is much larger than $a(i)$, the $s(i)$ value (silhouette coefficient) will be high. In the clustering on the right, $a(i)$ would be closer in value to $b(i)$, leading to a smaller value of $s(i)$. Though shown in two dimensions, the general concept applies to higher dimensional spaces.

With both of the above approaches, one would expect the following:

- The *rejection rate*, defined as the proportion of samples deemed unclassifiable, should increase as the thresholds are increased.
- The classification accuracy would increase when those samples deemed unclassifiable are excluded from the CA calculation, since the more questionable samples (i.e., lower confidence) are omitted.

4.3 Experiment #1 – Initial Feasibility Study

4.3.1 Experiment #1: Bacteria Sample Preparation

Two bacteria – *E. coli* DH5 α and *Listeria innocua* – were cultured onto agar plates (based on BHI broth for *E. coli* and LB broth for *Listeria*) using the conventional procedures (incubated at 37°C for 16h). This resulted in the appearance of distinct colonies (see Figure 4.4). Both plates were then refrigerated at 4°C to arrest further growth of the colonies.

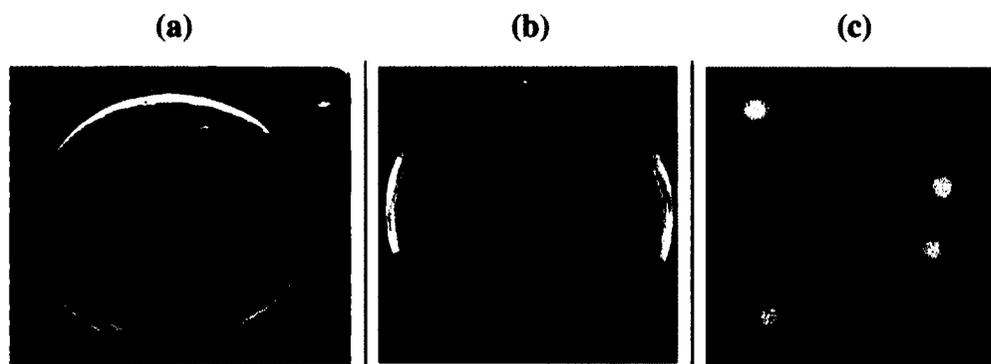


Figure 4.4: Photograph of bacteria colonies grown on agar substrate: a) *E. coli* DH5 α , b) *Listeria innocua*, c) enlarged view showing four *Listeria* colonies (for detail).

Each sample used for e-nose analysis consisted of a 10mL vial containing a single colony of bacteria suspended in 200 μ L of PBS. To prepare this sample, the colony was

removed by brushing lightly across the colony with a sterile pipette tip to collect as much of the colony as possible. Care was taken not to capture any agar substrate. The colony was then placed in a 10mL vial that had 200 μ L PBS previously added (as a suspension medium). Following this, the pipette tip was rinsed using five aspiration/ejection cycles to ensure that as much of the colony as possible was transferred to the vial. Immediately thereafter, the vials were sealed with PTFE-lined caps.

4.3.2 Experiment #1: Sampling Protocol

Samples were prepared and run on four separate days within a week-long period. On each day (except the first), there were 30 samples prepared – ten containing *E. coli* + PBS (Label: E), ten containing *Listeria* + PBS (Label: L), and ten containing 200 μ L PBS only (as a control, Label: P). On the first day, only 15 samples (five of each category) were prepared. A total of 105 samples were processed (see Table 4.2). After sample preparation, the vials were transferred to a cooling tray (maintained at 4°C to inhibit growth during the measurement sequence) mounted on the HS100 autosampler, and the e-nose sequence was started immediately thereafter. During the sequence, samples were processed in a cyclical manner (Vial 1: E, Vial 2: L, Vial 3: P, Vial 4: E, Vial 5: L, Vial 6: P, ...).

Sample Type	Category Label	Number of Samples
<i>E. coli</i> DH5 α	E	35
<i>Listeria innocua</i>	L	35
Phosphate buffered saline	P	35
	Total:	N = 105

Table 4.2: Summary of sample numbers and category labels used for Experiment #1.

4.3.3 Experiment #1: Results

4.3.3.1 Overall Results with All Samples Aggregated

In Figure 4.5, the results obtained for both unsupervised PCA (Figure 4.5(a)) and supervised ULDA (Figure 4.5(b)) applied to the feature vectors are illustrated (by plotting the first two components against one another). Each point in these plots represents a single bacteria sample in the dimensionality-reduced feature space. There is a clear improvement in separation between all three categories when ULDA is used (as compared to PCA). This observation is to be expected because no category information is used in PCA, while ULDA specifically looks for new features that separate the clusters based on labeled samples. The PCA analysis shows that the responses for the samples containing PBS only (no bacteria colony) have a degree of overlap with both E and L samples (which contain bacteria). This demonstrates that the PBS odour signature is only slightly modified by the addition of the bacteria colony. It requires the more powerful ULDA technique to emphasize the discriminating information (the signal being sought) from the noise, and the result is distinct clustering in the ULDA space.

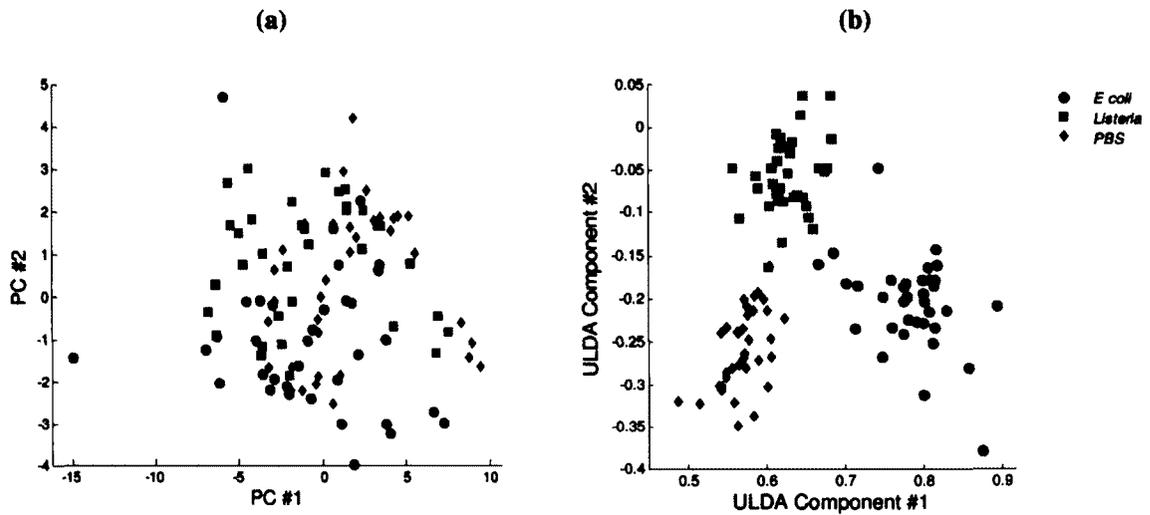


Figure 4.5: Experiment #1: Clustering results for all samples from all runs combined: a) PCA, b) ULDA.

Upon inspection of Table 4.3, it is seen that the maximum classification accuracy achieved is 92.4%. Also shown is the classifier performance when other sets of input features are used. The use of PCA with two components ($CA = 51.4\%$) is notably better than what would be expected from random guessing in this three-category problem (33.3%), and this improvement increases with the number of retained components [23]. ULDA provides the same classification performance (with only two components) as the entire original feature vector (with dimension = 24). The confusion matrix in Table 4.4 shows the distribution of classification errors.

Input Features	Category E	Category L	Category P	Overall Classification Accuracy, %
All 24 (no DR applied)	88.6	91.4	97.1	92.4
2 ULDA components	88.6	91.4	97.1	92.4
First 2 principal components	62.9	51.4	40.0	51.4
First 3 principal components	85.7	68.6	65.7	73.3

Table 4.3: Experiment #1: Overall and pre-category classification accuracy, for various input feature combinations.

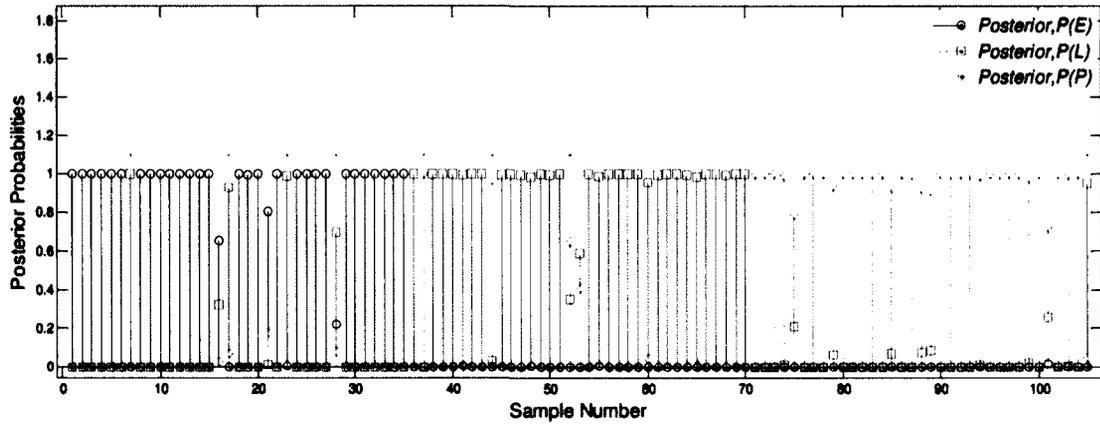
		CLASSIFICATION RESULT		
		E	L	P
TRUE CLASS	E	31	4	0
	L	0	32	3
	P	0	1	34

Table 4.4: Experiment #1: Confusion matrix for the three-category classification problem.

Finally, the problem was re-cast using only the two bacteria categories (E and L) by removing the PBS samples. Doing this gave a classification accuracy of 92.9% (using the LDA classifier with no dimensionality reduction), representing 65/70 correctly classified samples. With the samples used in this chapter (in which the *presence* of bacteria is inherently confirmed by virtue of having obtained a colony), this is a very relevant result (identification not detection).

The accuracies reported above are an encouraging result, but it is instructive to explore the confidence with which these classification decisions are made. Figure 4.6 illustrates both confidence measures (calculated for each sample) described in Section 4.2.5. The posterior values shown in Figure 4.6(a) are output by the LDA classifier during the LOO cross validation procedure. In this work, estimates of the priors $p(\omega_c)$ for $c = 1, \dots, C$ were based on the relative frequency of samples in the training set. For most samples, a single category gives a posterior probability very close to one (with the other two being far lower) – one would have a high confidence in these results. For other samples, two or more of the posterior probabilities are closer to each other, and the confidence measure should be penalized. In Figure 4.6, it is evident that the samples with very low classifier confidence are more likely to be misclassified (locations indicated with asterisks) than those with high confidence in the result.

(a)



(b)

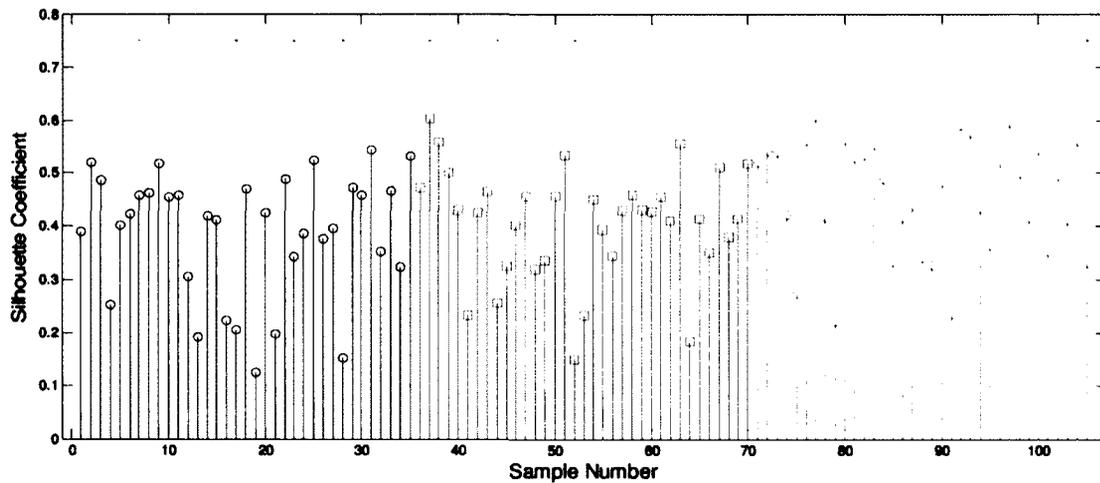


Figure 4.6: Experiment #1: Illustration of confidence measures for classifier results. Samples 1-35 are from category E, samples 36-70 are from category L, and samples 71-105 are from category P. In (a), the posterior probabilities of all three categories are plotted for each sample. In (b), the silhouette coefficient of each sample is shown. The locations of misclassified samples (as determined by the LOO procedure) appear as magenta asterisks at the top of these plots.

Figure 4.7 further illustrates the effectiveness of these confidence measures. Shown is the ULDA representation of all samples (same data as shown in Figure 4.5(b)) in which three arbitrarily selected samples have been highlighted. The first of these (sample #51, a *Listeria* sample) is found in the center of the L cluster so intuitively, one would expect the confidence measures for this sample to be high (relative to the others). Conversely,

the second sample (sample #16, an *E. coli* sample) is nestled between the E and L clusters so one would not have a high degree of confidence in the classifier output (regardless of the output category). This intuition is borne out with the calculated confidence measures:

- **Sample 51 (*Listeria*):** $p(\omega_E | \mathbf{f}_{51}) = 0$, $p(\omega_L | \mathbf{f}_{51}) = 1.00$, $p(\omega_P | \mathbf{f}_{51}) = 0$

$s(51) = 0.53$ (which is among the highest, as evident in Figure 4.6(b)). In this case, the classification was correct (with high confidence).

- **Sample 16 (*E. coli*):** $p(\omega_E | \mathbf{f}_{16}) = 0.65$, $p(\omega_L | \mathbf{f}_{16}) = 0.32$, $p(\omega_P | \mathbf{f}_{16}) = 0.03$

$s(16) = 0.22$ (which is among the lowest, as evident in Figure 4.6(b)). In this case, the classification was correct (with lower confidence).

The third data point (sample #7, an *E. coli* sample) is an example of one that was *incorrectly* classified (with high confidence):

- **Sample 7 (*E. coli*):** $p(\omega_E | \mathbf{f}_7) = 0$, $p(\omega_L | \mathbf{f}_7) = 1.00$, $p(\omega_P | \mathbf{f}_7) = 0$

$s(7) = 0.46$ (which is among the highest, as evident in Figure 4.6(b))

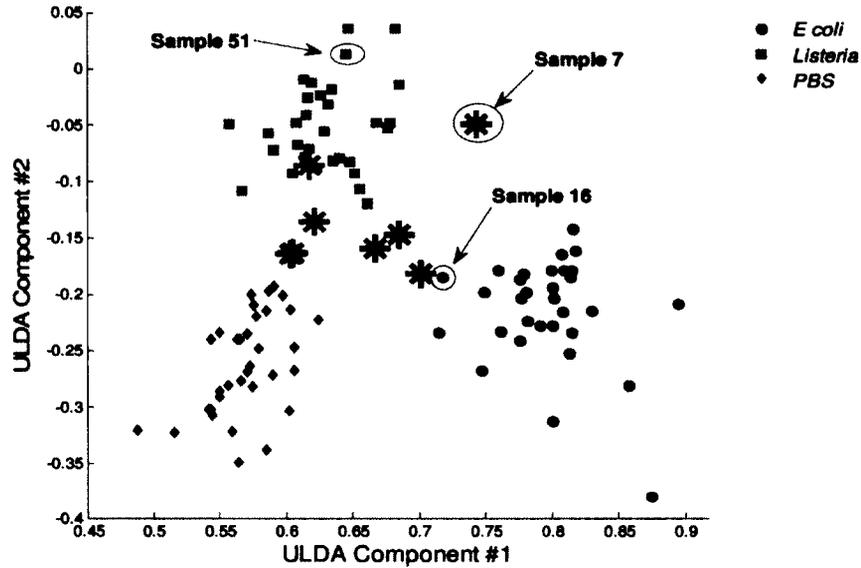


Figure 4.7: Experiment #1: Illustration of confidence measures for classifier results, specific samples. Sample 51 is entirely contained within the L cluster, while Sample 16 is closer to the boundaries between clusters. (The locations of misclassified samples, as determined by the LOO procedure, appear with large asterisks superimposed on the sample points. For these points, the asterisk colour represents the classifier result while the marker colour represents the true category).

The classification performance was assessed as the thresholds T_p and T_s were varied, with results shown in Figure 4.8. As expected, enforcing a higher degree of confidence in the classifier output (by increasing T_p and T_s) results in: a) a higher proportion of samples rejected as unclassifiable and b) higher classification accuracies among the remaining samples (since only the most confident results are included in the CA calculation). The location of an appropriate operating point (specified with a T_p or T_s constant) would represent a optimal trade-off between these two (as shown in Figure 4.8(a) with $T_p=0.7$). The representation on the right in this figure is called an *accuracy rejection curve* (ARC) [97].

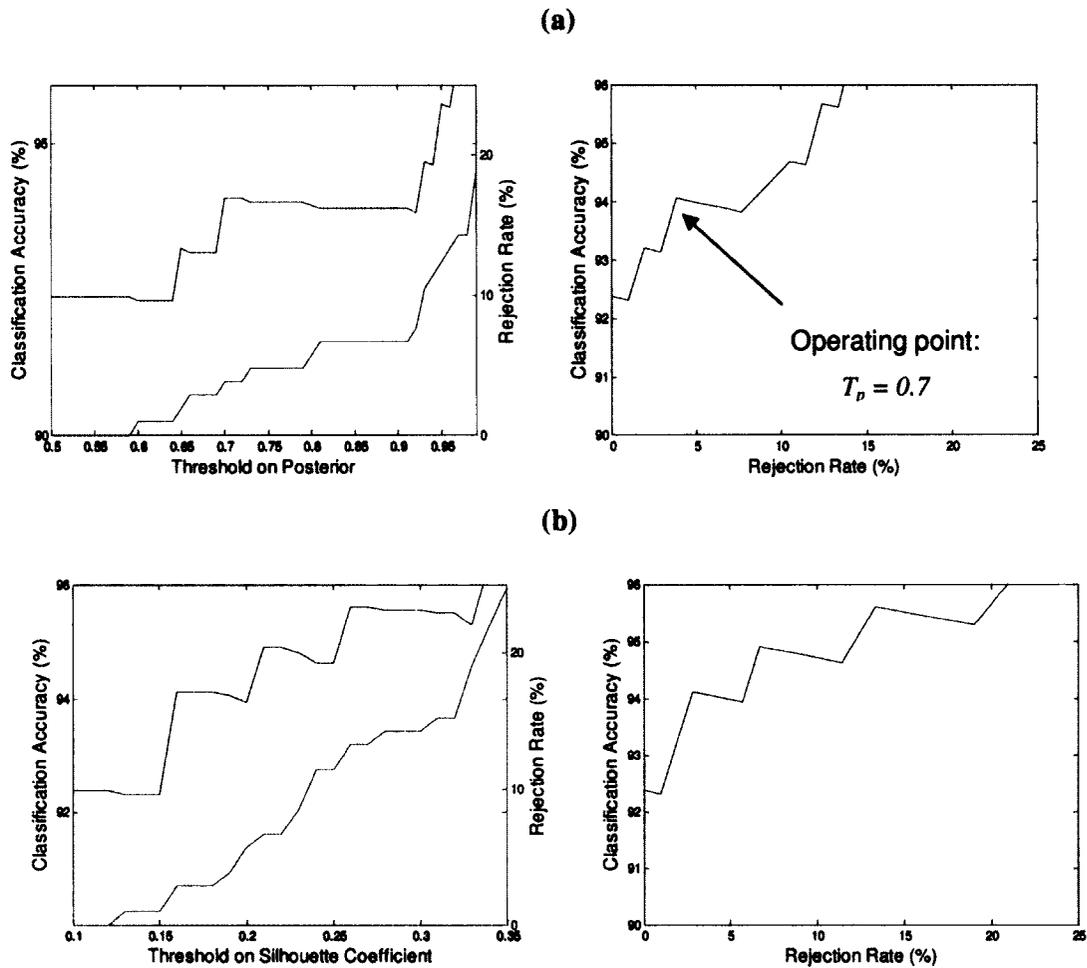


Figure 4.8: Experiment #1: Effect of varying confidence threshold on classification accuracy and rejection rate: (a) posterior probabilities and (b) silhouette coefficients. In the ARC on the right, each data point corresponds to a unique threshold value. A possible operating point is indicated on the posterior-based ARC in (a).

4.3.3.2 Results for Each Processing Day

Data collection for this experiment was performed on four separate days over a week-long period. Mindful of the possibility of a time-varying drift effect (described in Chapter 3), it was instructive to investigate the extent to which discrimination performance of the system varied during the week. In Figure 4.9, PCA results are shown (using the same transformation matrix as was used for Figure 4.5(a)). In this figure, however, the samples from each category – E, L, and P – have been further subdivided based on the day in

which the sample was processed. Within each category, the four clusters (defined by run number) do not appear to show a distinct trajectory in the PCA space. In Figure 4.10, the analogous ULDA results are shown. It is important to realize that the ULDA transformation matrix used in Figure 4.10 is the same as that used in Figure 4.5(b) – i.e. the grouping was performed using *three categories* (based on sample type – E, L, and P) and *not twelve categories* (based on sample type *and* run number). Again, the four clusters (defined by run number) for each bacteria type (and PBS) do not appear to show a distinct trajectory in the ULDA space. This implies the absence of any significant confounding drift-like effect in this measurement system – the system performance remains satisfactory over the duration of the experiment.

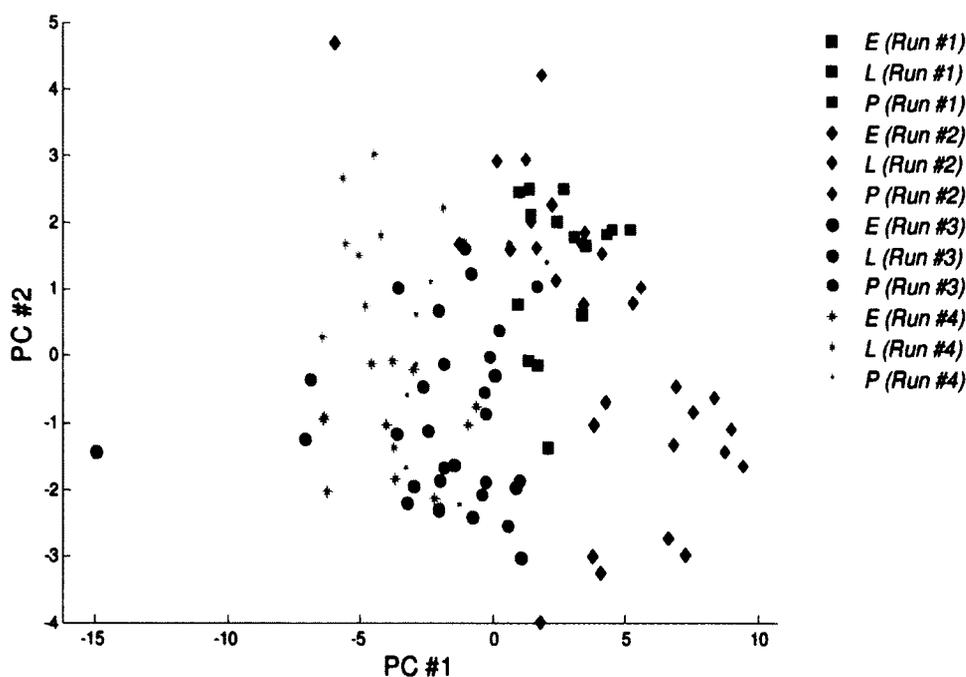


Figure 4.9: Experiment #1: PCA results with samples further categorized by the run number. Run #1 consisted of 15 samples (5 from each category), subsequent runs consisted of 30 samples (10 from each category). It is evident that as the experiment progressed, the E, L, and P categories do not exhibit a noticeable trajectory based on run number.

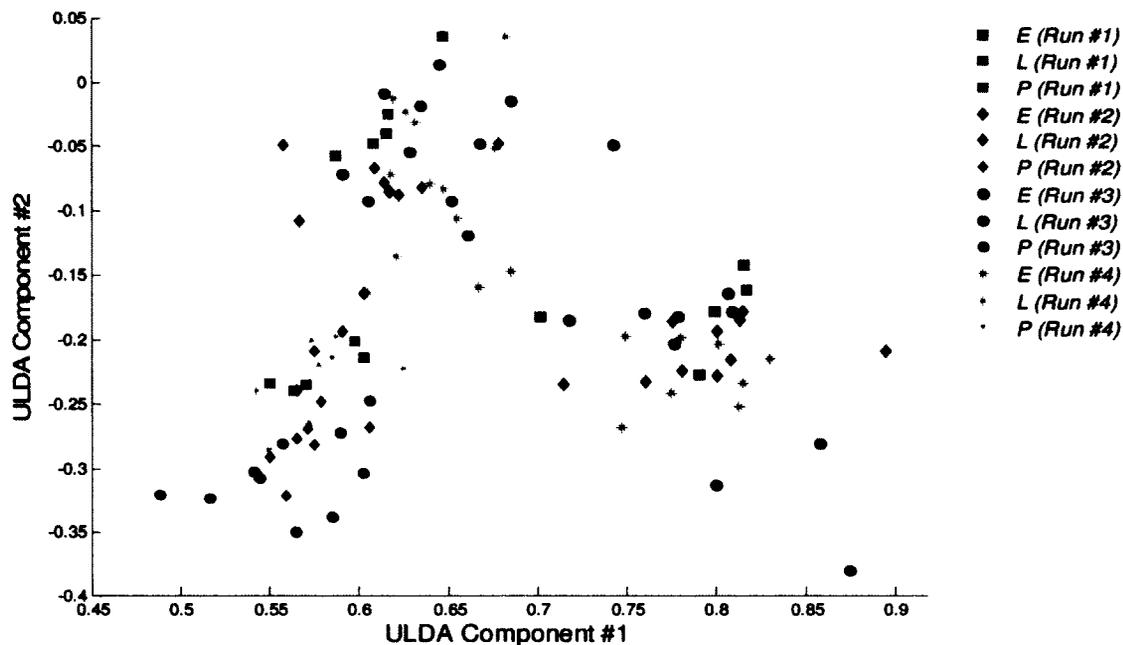


Figure 4.10: Experiment #1: ULDA results with samples further categorized by the run number. Run #1 consisted of 15 samples (5 from each category), subsequent runs consisted of 30 samples (10 from each category). It is evident that as the experiment progressed, the E, L, and P categories do not exhibit a noticeable trajectory based on run number. The ULDA transformation matrix is based on the three sample categories – E, L, and P.

4.3.4 Experiment #1: Conclusions and Discussion

The results shown above represent the first time that odour responses from individual bacteria colonies (as measured by a MOS e-nose) have been used to distinguish between bacteria types. Classification accuracies of greater than 90% were achieved (when employing supervised DR methods). Additionally, by avoiding the use of nutrient broth in the sample (instead using only a single colony and neutral PBS) it appears that the system performance should not be compromised by a time-varying drift (as was the case for the broth-based samples in Chapter 3).

Some comments regarding the colonies and the method used to sample them are warranted:

1. The colonies that grew on the plates had varied morphologies. The *Listeria* colonies were smaller and grew higher on the agar than did the *E. coli* colonies, which were generally wider and grew closer to the surface,
2. There was a natural variation in colony sizes for both bacteria (which overlapped), so sampling the bacteria from the plates was performed randomly using all available colony sizes. Doing this instilled some confidence that the inter-species separation demonstrated is not primarily due to sample volume/size,
3. When removing the colony with the pipette tip, it was not possible to transfer the entire colony to the vial (a small amount of residual bacteria remained on the agar surface), and
4. Despite careful removal of bacteria, it is likely that in some cases, a small amount of residual agar substrate was transferred to the vial.

The results obtained in Experiment #1 illustrate that the method is quite robust despite these sampling variations.

4.4 Experiment #2 – Validation Study

4.4.1 Experiment #2: Bacteria Sample Preparation

Experiment #2 was designed to investigate whether the odour discrimination ability demonstrated in Experiment #1 could be seen when colonies from more than two bacteria types were included in the training set, while simultaneously controlling for the factors pointed out above (varying colony size and different agar growth media).

To do this, four bacteria were considered – *E. coli* DH5 α , *Listeria innocua* (both previously considered in Experiment #1) and two additional non-pathogenic bacteria,

Enterococcus faecalis and *E. coli* Biotype I. All were cultured onto agar plates but unlike Experiment #1, the same agar growth medium (BHI) was used for all four bacteria types. Also in contrast to Experiment #1, the incubation time was adjusted (between 8h-24h) to ensure the emergence of colonies of roughly similar size across all bacterial species (the *E. coli* bacteria grow preferentially in the presence of BHI so these did not require as long an incubation time as the other bacteria to reach the desired size). As in Experiment #1, incubation of colonies occurred at a temperature of 37°C. Sample preparation proceeded exactly as described previously in Section 4.3.1 (each sample vial contained a single colony in suspension with 200µL of PBS).

4.4.2 Experiment #2: Sampling Protocol

In total, 30 vials were prepared (representing six vials from each of the five categories – *E. coli* DH5α (Label: E), *Listeria innocua* (Label: L), *Enterococcus faecalis* (Label: F), and *E. coli* Biotype I (Label: I), and PBS control (Label: P). The vials were then immediately transferred to the 4°C cooling tray to arrest growth that might take place during the measurement sequence. E-nose instruments are known to suffer from a sensor “memory” effect – when sample types are changed on consecutive sniffs, there can be incomplete purging of the sensor surfaces and subsequent samples may have temporarily altered odour profiles [83]. In addition, previous work by Trincavelli *et al.* [95] on blood cultures has suggested that better classification performance can be achieved by combining the information available from multiple repeated e-nose responses on the same sample. Unlike Experiment #1, then, *five consecutive e-nose responses (sniffs) were taken from the same vial.* To accomplish this, sniffs were performed by the syringe re-

puncturing the same vial five times. Prior to each sniff, the headspace generation (see Table 4.1) was re-performed on the same vial. All bacteria types were then processed in a cyclical manner using the same method. Also in contrast to Experiment #1 (where e-nose responses were recorded for four separate trays of samples on different days), the 30 vials for Experiment #2 were prepared at the beginning of the experiment and all e-nose responses (including multiple sniffs from each vial) were recorded in a *single run* lasting approximately 56h. There were 30 sniffs (6 vials of each x 5 sniffs/vial), for each of the 5 sample types – E, F, I, L, and P – representing 150 in total. Table 4.5 makes explicit the method in which the e-nose sampling occurred.

Sniff Number (Chronological)	Vial Number and Contents
1	Vial #1: P
2	Vial #1: P
3	Vial #1: P
4	Vial #1: P
5	Vial #1: P
6-10	Vial #2: E
11-15	Vial #3: L
16-20	Vial #4: I
21-25	Vial #5: F
26-30	Vial #6: P
31-35	Vial #7: E
36-40	Vial #8: L
41-45	Vial #9: I
46-50	Vial #10: F
...	...
126-130	Vial #26: P
131-135	Vial #27: E
136-140	Vial #28: L
141-145	Vial #29: I
146-150	Vial #30: F
Total Sniffs: 150	Total Vials: 30

Table 4.5: Sampling protocol for Experiment #2 illustrating the order in which consecutive e-nose responses were recorded. A total of 150 sniffs were recorded from 30 vials.

4.4.3 Experiment #2: Results

4.4.3.1 Overall Results with All Samples Aggregated

When all 150 e-nose responses are aggregated into a single data set, the results are shown in Figure 4.11(a), for PCA, and Figure 4.11(b), for ULDA, using the first two components in each.

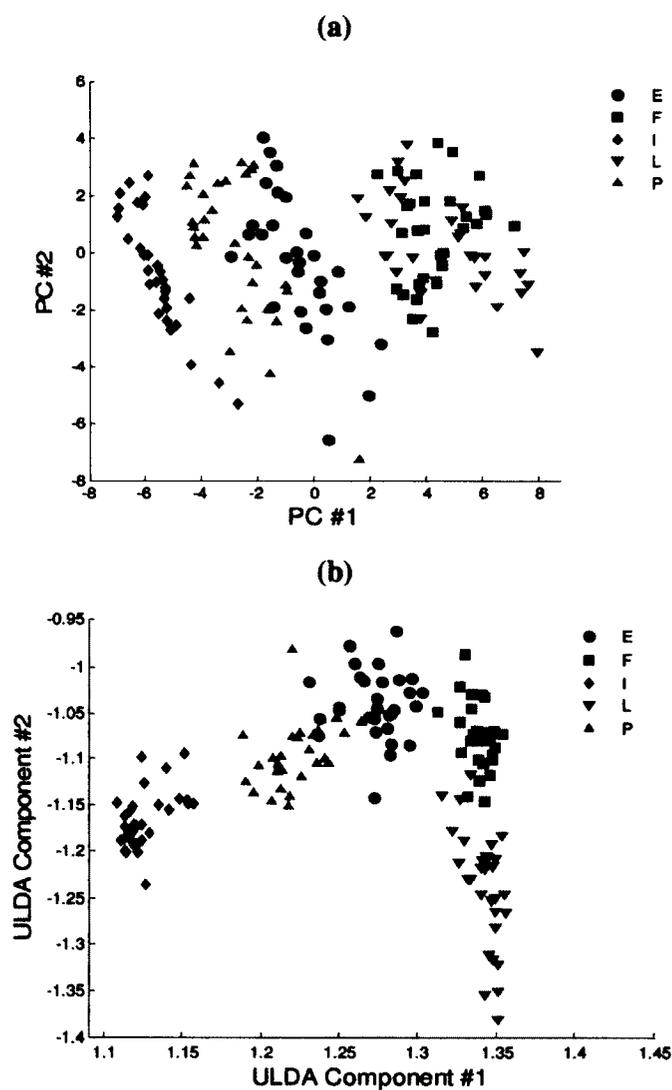


Figure 4.11: Experiment #2: Clustering results with all 150 sniffs (from 30 bacteria samples) combined: a) PCA, b) ULDA.

Like elsewhere in the thesis, a LOO cross validation (in conjunction with an LDA classifier) was used to estimate classification performance. However, there is an important *caveat* in interpreting the results presented above. In this experiment, the LOO approach withheld a single sniff (the test set), while training on the other $N - 1 = 149$ sniffs. Due to the manner in which the sampling was performed (5 sniffs/vial), four of those 149 training sniffs were taken from the same vial that the test sniff was taken from. Table 4.6 provides the classification results and confusion matrix when this LOO method is used. (In this case, all five ULDA features were input to the classifier). The overall classification accuracy achieved was 92%.

		CLASSIFICATION RESULT				
		E	F	I	L	P
TRUE CLASS	E					5
	F					
	I					
	L	1	3		26	
	P	3				27

Table 4.6: Experiment #2: Confusion matrix for the five-category classification problem using LOO (i.e. test set size = one sample, training set size=149 samples). Blank entries represent zero. 5 ULDA features were used for classification.

In order to eliminate *any association* between the test set and the training set, a modified LOO method was used. In this scheme, the other four sniffs that originated from the same vial as the test sniff are eliminated from the training set prior to classification. This means that for each iteration of the LOO, there is a single test sniff and $(N - 1 - 4) = 145$ training sniffs. The overall classification accuracy achieved was 80.7% (again using all five ULDA features as input to the classifier). As would be expected, this method gave a result somewhat lower than above (this can be attributed to the isolation between test and training data in the modified LOO procedure, which was not strictly the case in

the LOO procedure described in the previous paragraph). Nevertheless, a CA value of over 80% is a considerable improvement over the accuracy expected by random guessing in this five category problem (20%). Table 4.7 shows the resulting confusion matrix for the modified LOO procedure.

		CLASSIFICATION RESULT				
		E	F	I	L	P
TRUE CLASS	E		2		5	7
	F	1				
	I					1
	L	1	4		25	
	P	8				22

Table 4.7: Experiment #2: Confusion matrix for the five-category classification problem using *modified* LOO (i.e. test set size = one sample, training set size=145 samples). Blank entries represent zero. 5 ULDA features were used for classification.

Finally (as was done in Experiment #1), the classification problem was then re-cast by omitting the sniffs from vials containing only PBS ($N = 30$) and retaining only those from the four bacteria types ($N = 120$). Again, recall that the *presence* of bacteria is intrinsically confirmed by the sampling method (having obtained a colony). This four-category identification problem yielded classification accuracies of 98.3% and 97.5% (using the LDA classifier and LOO and modified LOO methods described above, respectively, with five ULDA features as input to the classifier).

Figure 4.12 shows the effect of thresholding the posterior- and silhouette coefficient-based confidence measures on the classification accuracy of the system. The same general trend as was evident in the three-category case (Figure 4.8) is seen here (increased CA as the rejection rate increases). The ARC curves (also shown) would help determine which threshold represents the optimal tradeoff between the two.

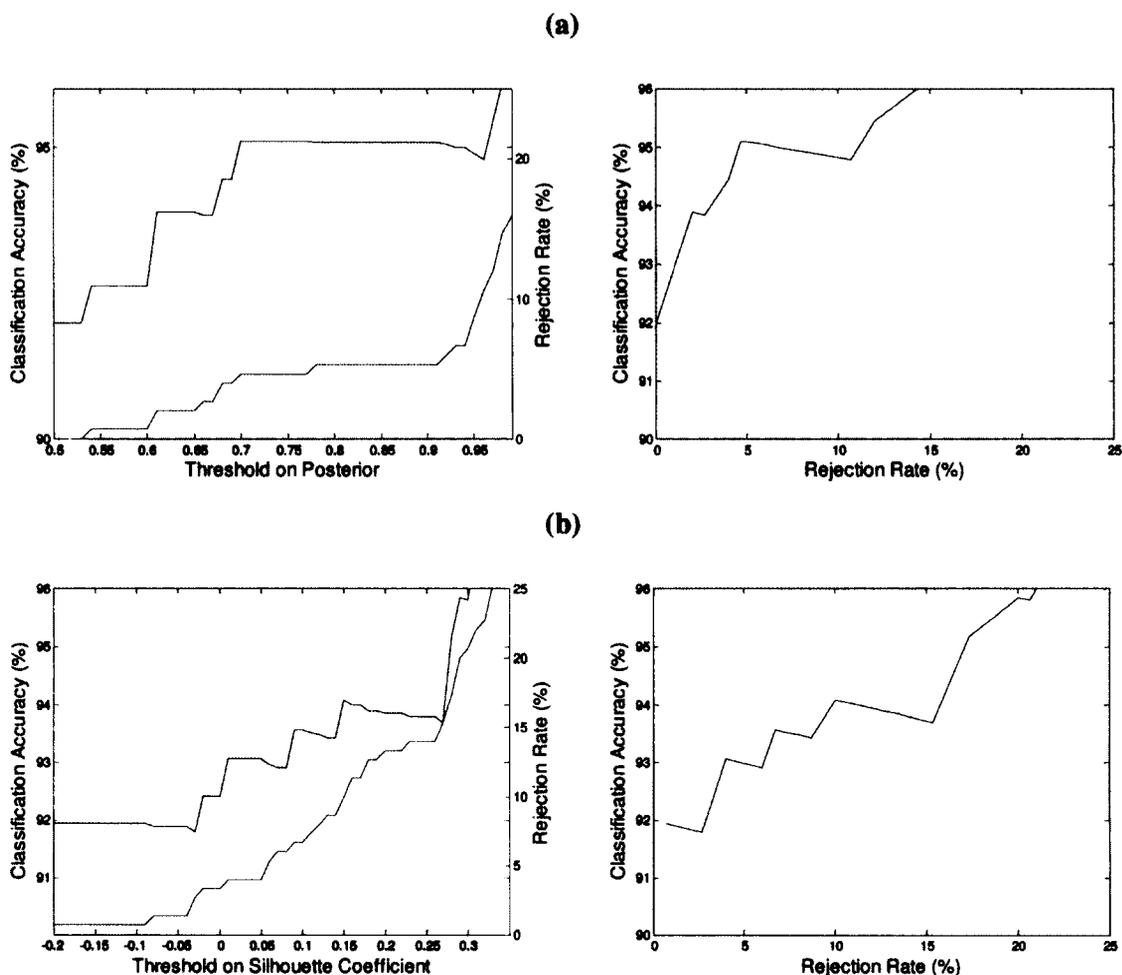


Figure 4.12: Experiment #2: Effect of varying confidence threshold on classification accuracy and rejection rate: (a) posterior probabilities and (b) silhouette coefficients. In the accuracy rejection curve on the right, each data point corresponds to a unique threshold value.

4.4.3.2 Per-Sniff Results

It is important to bear in mind that the results above included five separate e-nose responses (i.e. sniffs) from each vial (due to re-sampling from the same vial). It is instructive to investigate the discrimination performance that can be achieved when only one sniff per vial (instead of five) is considered, as demonstrated in Figure 4.13 (overall and per-category). In this case, all ULDA features were used to train an LDA classifier

and the LOO cross-validation method used 29 training samples and a single test sample (at each iteration).

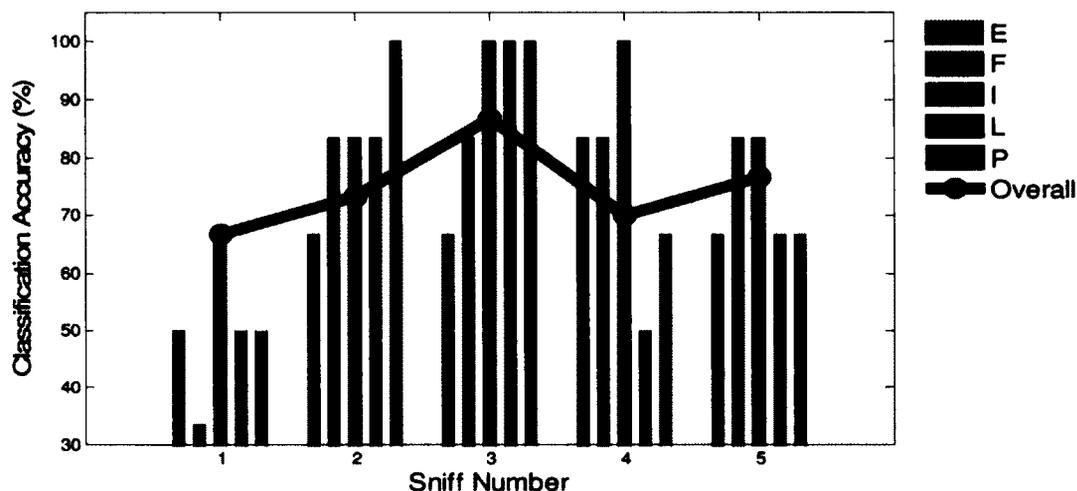


Figure 4.13: Experiment #2: Overall (solid black line) and per-category (colored bars) classification accuracy as a function of chronological sniff number. In this case, the feature vector consisted of the $f_{j,max}$ components only and classification used all ULDA features with LOO cross-validation.

The classification performance, as a function of sniff number, increases from 66.7% (using Sniff #1) to 86.7% (using Sniff #3) and drops off after this. Table 4.8 shows the confusion matrices that result from each sniff. Because Sniff #1 yielded the lowest accuracy, it is useful to examine the extracted features (as a function of sniff number) in an attempt to understand this effect. Recall that the Sniff #1 e-nose responses are collected immediately after five consecutive sniffs of the previous bacteria type has been processed (see Table 4.5). The decreased performance for Sniff #1 is likely due to the aforementioned sensor “memory” effect (see Section 4.4.2). Figure 4.14 illustrates this effect for two arbitrarily selected sensors for several transitions between vials. It is easily seen that the first sniff for each vial is notably different (generally a lower feature value) than the subsequent sniffs, suggesting that this sniff should be used for conditioning the sensors only (and discarded from subsequent analyses).

<p>Sniff #1 Overall CA = 66.7%</p> <table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="5">CLASSIFICATION RESULT</th> </tr> <tr> <th>E</th> <th>F</th> <th>I</th> <th>L</th> <th>P</th> </tr> </thead> <tbody> <tr> <th rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">TRUE CLASS</th> <th>E</th> <td style="background-color: black;"></td> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td style="background-color: black;"></td> <td></td> <td>2</td> <td></td> </tr> <tr> <th>I</th> <td></td> <td></td> <td style="background-color: black;"></td> <td></td> <td>2</td> </tr> <tr> <th>L</th> <td></td> <td>1</td> <td></td> <td style="background-color: black;"></td> <td>5</td> </tr> <tr> <th>P</th> <td>1</td> <td></td> <td>1</td> <td></td> <td style="background-color: black;"></td> <td>4</td> </tr> <tr> <th></th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>			CLASSIFICATION RESULT					E	F	I	L	P	TRUE CLASS	E		1		1	1	F				2		I					2	L		1			5	P	1		1			4								<p>Sniff #2 Overall CA = 73.3%</p> <table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="5">CLASSIFICATION RESULT</th> </tr> <tr> <th>E</th> <th>F</th> <th>I</th> <th>L</th> <th>P</th> </tr> </thead> <tbody> <tr> <th rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">TRUE CLASS</th> <th>E</th> <td style="background-color: black;"></td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td>1</td> <td style="background-color: black;"></td> <td></td> <td></td> <td></td> </tr> <tr> <th>I</th> <td></td> <td></td> <td style="background-color: black;"></td> <td></td> <td>1</td> </tr> <tr> <th>L</th> <td></td> <td>1</td> <td></td> <td style="background-color: black;"></td> <td>5</td> </tr> <tr> <th>P</th> <td>1</td> <td></td> <td></td> <td></td> <td style="background-color: black;"></td> <td>5</td> </tr> <tr> <th></th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>			CLASSIFICATION RESULT					E	F	I	L	P	TRUE CLASS	E		1	1	1	1	F	1					I					1	L		1			5	P	1					5							
			CLASSIFICATION RESULT																																																																																																				
		E	F	I	L	P																																																																																																	
TRUE CLASS	E		1		1	1																																																																																																	
	F				2																																																																																																		
	I					2																																																																																																	
	L		1			5																																																																																																	
	P	1		1			4																																																																																																
		CLASSIFICATION RESULT																																																																																																					
		E	F	I	L	P																																																																																																	
TRUE CLASS	E		1	1	1	1																																																																																																	
	F	1																																																																																																					
	I					1																																																																																																	
	L		1			5																																																																																																	
	P	1					5																																																																																																
<p>Sniff #3 Overall CA = 86.7%</p> <table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="5">CLASSIFICATION RESULT</th> </tr> <tr> <th>E</th> <th>F</th> <th>I</th> <th>L</th> <th>P</th> </tr> </thead> <tbody> <tr> <th rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">TRUE CLASS</th> <th>E</th> <td style="background-color: black;"></td> <td></td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td style="background-color: black;"></td> <td></td> <td>1</td> <td></td> </tr> <tr> <th>I</th> <td></td> <td></td> <td style="background-color: black;"></td> <td></td> <td></td> </tr> <tr> <th>L</th> <td></td> <td></td> <td></td> <td style="background-color: black;"></td> <td>6</td> </tr> <tr> <th>P</th> <td>1</td> <td></td> <td></td> <td></td> <td style="background-color: black;"></td> <td>5</td> </tr> <tr> <th></th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>			CLASSIFICATION RESULT					E	F	I	L	P	TRUE CLASS	E				1	1	F				1		I						L					6	P	1					5								<p>Sniff #4 Overall CA = 70%</p> <table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="5">CLASSIFICATION RESULT</th> </tr> <tr> <th>E</th> <th>F</th> <th>I</th> <th>L</th> <th>P</th> </tr> </thead> <tbody> <tr> <th rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">TRUE CLASS</th> <th>E</th> <td style="background-color: black;"></td> <td></td> <td></td> <td>1</td> <td>2</td> </tr> <tr> <th>F</th> <td>1</td> <td style="background-color: black;"></td> <td></td> <td></td> <td></td> </tr> <tr> <th>I</th> <td></td> <td></td> <td style="background-color: black;"></td> <td></td> <td></td> </tr> <tr> <th>L</th> <td>1</td> <td>2</td> <td></td> <td style="background-color: black;"></td> <td>3</td> </tr> <tr> <th>P</th> <td>2</td> <td></td> <td></td> <td></td> <td style="background-color: black;"></td> <td>4</td> </tr> <tr> <th></th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>			CLASSIFICATION RESULT					E	F	I	L	P	TRUE CLASS	E				1	2	F	1					I						L	1	2			3	P	2					4							
			CLASSIFICATION RESULT																																																																																																				
		E	F	I	L	P																																																																																																	
TRUE CLASS	E				1	1																																																																																																	
	F				1																																																																																																		
	I																																																																																																						
	L					6																																																																																																	
	P	1					5																																																																																																
		CLASSIFICATION RESULT																																																																																																					
		E	F	I	L	P																																																																																																	
TRUE CLASS	E				1	2																																																																																																	
	F	1																																																																																																					
	I																																																																																																						
	L	1	2			3																																																																																																	
	P	2					4																																																																																																
<p>Sniff #5 Overall CA = 76.7%</p> <table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="5">CLASSIFICATION RESULT</th> </tr> <tr> <th>E</th> <th>F</th> <th>I</th> <th>L</th> <th>P</th> </tr> </thead> <tbody> <tr> <th rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">TRUE CLASS</th> <th>E</th> <td style="background-color: black;"></td> <td></td> <td></td> <td>1</td> <td>2</td> </tr> <tr> <th>F</th> <td></td> <td style="background-color: black;"></td> <td></td> <td></td> <td></td> </tr> <tr> <th>I</th> <td></td> <td></td> <td style="background-color: black;"></td> <td></td> <td></td> </tr> <tr> <th>L</th> <td>1</td> <td>1</td> <td></td> <td style="background-color: black;"></td> <td>4</td> </tr> <tr> <th>P</th> <td>2</td> <td></td> <td></td> <td></td> <td style="background-color: black;"></td> <td>4</td> </tr> <tr> <th></th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>				CLASSIFICATION RESULT					E	F	I	L	P	TRUE CLASS	E				1	2	F						I						L	1	1			4	P	2					4																																																										
				CLASSIFICATION RESULT																																																																																																			
		E	F	I	L	P																																																																																																	
TRUE CLASS	E				1	2																																																																																																	
	F																																																																																																						
	I																																																																																																						
	L	1	1			4																																																																																																	
	P	2					4																																																																																																

Table 4.8: Experiment #2: Confusion matrices for the five category classification (presented for each sniff). Blank entries represent zero. All ULDA features were used as inputs to the classifier.

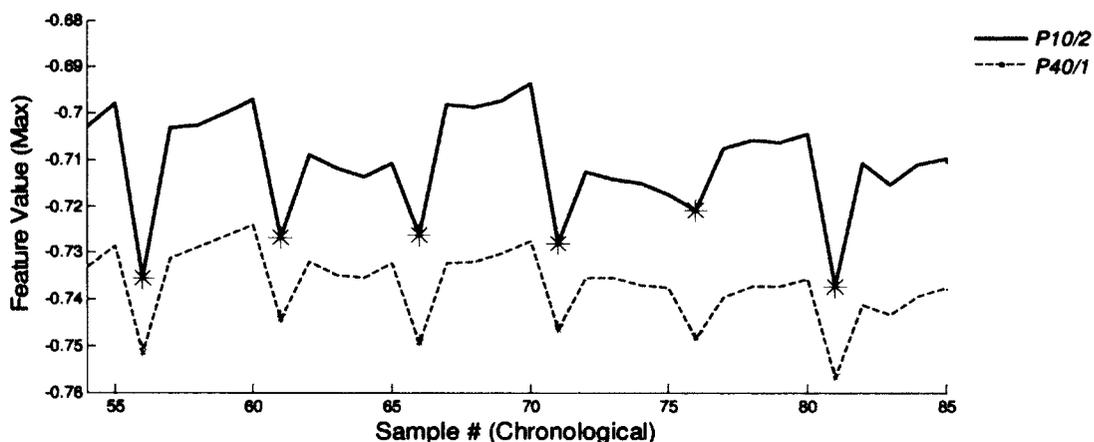


Figure 4.14: Experiment #2: Features extracted from consecutive sniffs demonstrating the significant difference in response values for the first e-nose response measured from each vial (denoted with an asterisk) for two MOS sensors (P10/2 and P40/1). As indicated in the figure: Sniffs 56-60 are from Vial 3 (Category F), Sniffs 61-65 are from Vial 3 (Category I), etc.

An effect similar to this was seen in previous work [95]. As in this case, the sensors' conditioning effect provides a likely explanation for the lower accuracy achieved using only the Sniff #1 samples.

4.4.3.3 Results from Combining Sniffs

The previous section demonstrated that the system's ability to discriminate between bacteria samples varied based on the position in the consecutive sampling protocol (i.e. sniff number). This section investigates whether increased classification performance can be achieved by creating a *composite* e-nose response (using the five consecutive sniffs available) for each vial of bacteria. The work described here is distinct from Section 4.4.3.1 (which presented aggregate results using all 150 e-nose responses).

In a classification system, fusion can occur on multiple levels (either alone or in combination), including:

- sensor modality level (in which different sensor types are combined in forming the features, e.g., in [101] where both e-nose and electronic tongue sensors are employed),
- data level (in which the feature vectors from a single modality are combined in some manner), and
- classifier level (in which the opinion of individual classifier experts are combined) [102].

The technique below describes a straightforward approach where fusion is performed at the *data level*. For each vial, composite e-nose responses for each vial i were calculated by computing the mean of feature values for various subsets S (of size n_S) of the five available sniffs:

$$\mathbf{f}_{comp,i} = \frac{1}{n_S} \left[\sum_{s \in S} f_{1,s} \quad \sum_{s \in S} f_{2,s} \quad \cdots \quad \sum_{s \in S} f_{12,s} \right] \quad (4.6)$$

where $f_{j,s}$ represents the value of feature j ($j = 1, 2, \dots, 12$) during sniff s . (Only the $f_{j,max}$ feature was used in this analysis). This was done for each of the thirty vials, and LOO cross validation was used to calculate *CA* (i.e. the same composite sniff combinations were used for *both* training and test sets). All five ULDA features were used as inputs to the classifier. Table 4.9 illustrates the results for various subsets S with the resulting confusion matrices shown in Table 4.10.

Composite subset, S	Classification Accuracy (%)
{1,2,3,4,5}	80
{2,3,4,5}	96.7
{3,4,5}	90
{2,3,4}	90
{2,3}	90
{4,5}	83.3
{1}	66.7
{2}	73.3
{3}	86.7
{4}	70
{5}	76.7

Table 4.9: Classification accuracies using composite e-nose responses using various subsets of sniffs. In all cases, the composite was formed by averaging the feature responses. Strictly, the last four rows are not composites (from single sniffs), but the performance is duplicated here to facilitate the comparison (data also appears in Figure 4.13 – overall results, solid black line – and Table 4.8). All ULDA features were used as inputs to the classifier.

<p style="text-align: center;">$S = \{1,2,3,4,5\}$ Overall CA = 80%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E				1	2
	F					
	I					
	L					5
	P	2				

<p style="text-align: center;">$S = \{2,3,4,5\}$ Overall CA = 96.7%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E					
	F					
	I					
	L					6
	P	1				

<p style="text-align: center;">$S = \{3,4,5\}$ Overall CA = 90%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E					2
	F					
	I					
	L					6
	P	1				

<p style="text-align: center;">$S = \{2,3,4\}$ Overall CA = 90%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E				1	2
	F					
	I					
	L					6
	P					

<p style="text-align: center;">$S = \{2,3\}$ Overall CA = 90%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E				1	1
	F					
	I					1
	L					6
	P					

<p style="text-align: center;">$S = \{4,5\}$ Overall CA = 83.3%</p>						
		CLASSIFICATION RESULT				
TRUE CLASS		E	F	I	L	P
	E					1
	F	1				
	I					
	L		1			4
	P	1				

Table 4.10: Experiment #2: Confusion matrices for the five category classification (presented for each of the subsets S used in the composite sniff). Confusion matrices for single sniffs were shown in Table 4.8. Blank entries represent zero. All ULDA features were used as inputs to the classifier.

4.4.4 Experiment #2: Conclusions and Discussion

It is evident in Figure 4.11 and Tables 4.6 and 4.7 that the ability for the e-nose to discriminate between the odours of single bacteria colonies (initially demonstrated in Experiment #1) has also been validated in Experiment #2. This validation was performed by: a) controlling for the colony sizes, b) controlling for the agar substrate type and c) performing the analysis with a larger range of bacteria (four instead of two). The use of supervised DR resulted in separated clusters for each bacteria type, and these results are consistent when both individual sniffs are considered as well as when all e-nose responses are aggregated. Using a similar classification method as was used in Experiment #1, CA values of over 90% (using LOO) and over 80% (using modified LOO) were obtained when all sniffs were aggregated. Using individual sniffs (one per sample vial), CA values in excess of 85% were reported. From the confusion matrices in Tables 4.6, 4.7, 4.8 and 4.10, it is worth noting that the samples in categories I (*E. coli* Biotype I) and F (*Enterococcus faecalis*) demonstrated close to perfect classification. Classification errors were spread more or less evenly over the other three categories. Two of the bacteria categories – E and I – are different strains from the same species (*Escherichia coli*), so it was reasonable to expect that their odour signatures might be very similar. This did not end up being the case, however – it is evident in the PCA results of Figure 4.11(a) that these categories' samples do not exhibit a high degree of overlap. Furthermore, inspection of the confusion matrices mentioned above revealed that sniffs of the category E (*E. coli* DH5 α) were very infrequently misclassified as belonging to category I (*E. coli* Biotype I), and vice versa. Due to the empirical nature of the MOS e-nose responses, it is very difficult to ascertain the reason for this behaviour. (Doing so

would require characterization of the VOCs resulting from the metabolic products of both *E. coli* strains with analytical techniques (such as GC/MS).

Moreover, by collecting multiple consecutive sniffs from each colony, it was possible to increase the classification accuracy of the system. First, it was demonstrated that the initial sniff was notably different from the others (presumably due to a sensor conditioning effect) and did not have the same ability to differentiate between bacteria types. Second, a composite e-nose response was formed by averaging various combinations of the remaining sniffs, and this approach was shown to yield CA values of over 95% (when the first sniff was omitted from the composite). These results are in agreement with the work of Trincavelli et al. [95], who processed blood cultures containing ten bacteria types repeatedly with an e-nose. In their work, however, the mechanism of ensembling between consecutive responses was performed at the classifier level using support vector machines. Indeed, the formulation of the composite sniff used in this chapter (essentially a mean calculation, while omitting the first sniff) is a very straightforward approach. While it proved effective, this could likely be improved upon with the application of other more sophisticated techniques available in the sensor fusion and classifier fusion literature that might be considered at multiple fusion levels [102].

Referring to Figure 4.13, it is unclear why the CA decreases after Sniff #3. Recall that each of the consecutive e-nose responses measured from a single vial employed the same headspace generation technique prior to sampling (e.g., heating of the vial to 50°C and agitation for 300s – see Table 4.1). It is suspected that by the later sniffs, the concentration of VOCs in the sample headspace has been exhausted, and this effect results in the decreased discrimination power. This can be ascertained with further work

in this area, which might investigate: a) the optimal number of consecutive sniffs to be taken from each vial, b) the optimal time delay between sniffs that would allow the headspace to recover (the delay between consecutive sniffs in the experiment above was roughly 22.5 minutes), and c) possible adjustment of headspace generation parameters to make a greater number of usable e-nose responses available to the composite measure.

Finally, it is worth highlighting that the promising classification accuracy reported in this chapter was achieved with “general purpose” MOS sensors (see Section 3.2.2.1). Future consultation with vendors of e-nose instruments might result in an optimized sensor selection better tuned to provide even better discrimination based on the VOCs emanating from bacteria samples.

4.5 Overall Summary and Conclusions

This chapter described the investigation of a novel sampling method (using a single bacterial colony in PBS suspension) and assessed the extent to which volatile patterns arising from this type of sample (as measured with a MOS e-nose) are sufficient to distinguish between four bacteria types (including two strains within the species *Escherichia coli*). Classification performance was shown to increase when multiple consecutive sniffs from the same bacteria were combined into a composite e-nose response. There are several papers that demonstrate that e-nose systems can distinguish bacteria based on *numerous* bacterial colonies grown on agar on a Petri dish (as opposed to a single colony). In these studies, the measurement is done by sniffing the entire Petri dish (containing pure cultures) [66], [94], [103]. The method described in this chapter holds a couple of important practical advantages over this type of sampling. Firstly, there

is no need for the Petri dish sampling apparatus. Secondly, when a blended/stomached food sample is streaked onto agar substrate, it is likely that colonies from a *mixture of bacteria* will appear (reflecting the diverse range of contaminants that arise in the field [38] – see Figure 4.1). The performance of a system trained to recognize odours based on bacteria of a single origin will likely be suspect when the incoming test sample contains a variety of bacteria. This becomes less of a concern with the sampling scheme described herein because the e-nose response is measured on a single colony (after removal from the plate).

Identification of bacteria based on single colonies would be of interest in food safety applications, as it is available relatively early in the inspection process (and might thereby generate faster overall results with decreased cost). Figure 4.15 highlights the stage at which this sample type is available. Using these samples, over 90% classification accuracy was achieved when considering four bacterial species – this degree of performance was a very encouraging result for an initial investigation. The method presented in this chapter could in the future comprise a pre-screening test, working together with the existing conventional methods. With this approach, determination of the presence and/or category of contaminants might be done with confidence using the proposed method for certain samples, while others would be deferred to the more sensitive tests for a definitive result. This chapter investigated two confidence measures – based on posterior probabilities and silhouette coefficients – that might prove to be of value in governing this decision. In the food safety application (where an incorrect result could have impact on human health), a conservative approach is merited. It is preferable for a classification system to “recognize its limits” by deferring the unclassifiable

samples to the slower but more accurate conventional methods (as opposed to forcing a measurement into a perhaps unsuitable category).

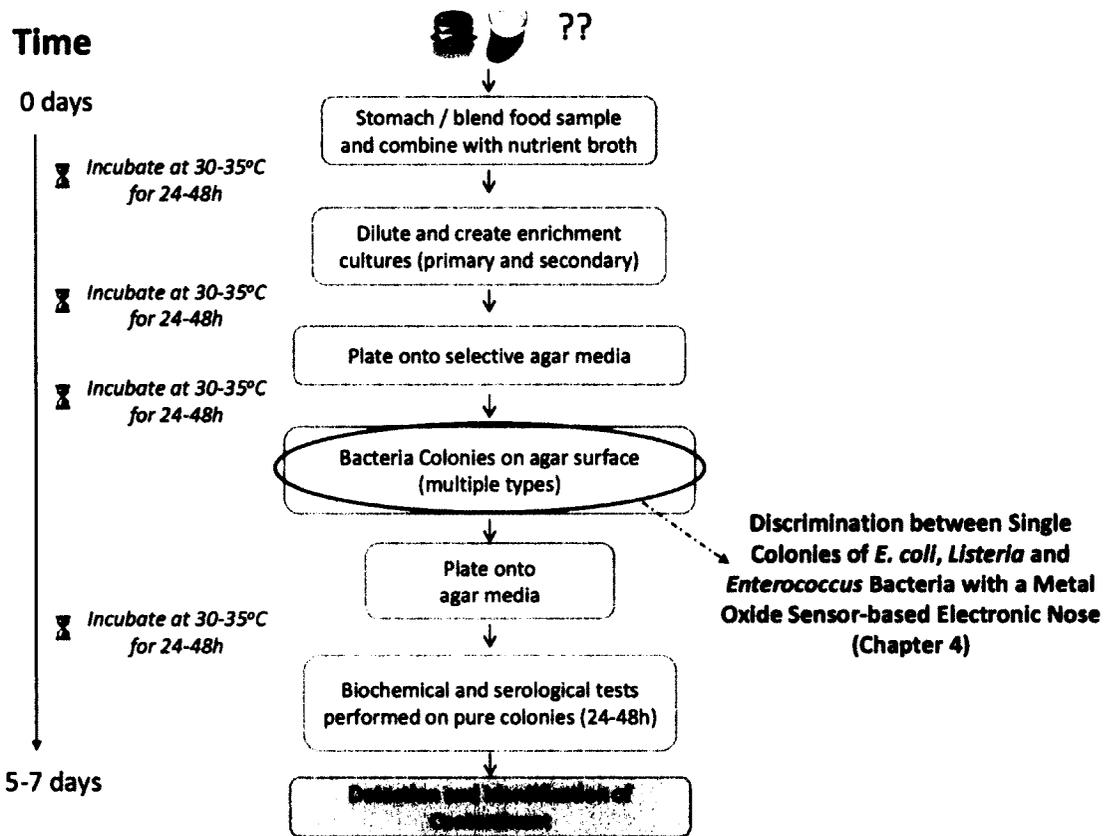


Figure 4.15: Simplified view of conventional methods. Highlighted block indicates the stage at which the sample type considered in Chapter 4 is available.

Chapter Five

Discrimination between *Listeria* Species with Surface-Enhanced Raman Spectroscopy

This chapter is concerned with discrimination between bacteria species (with particular attention to the genus *Listeria*). This is an important problem because, in general, not all species within a genus are harmful to human health. With *Listeria* for example, it is only *L. monocytogenes* that is considered to be an important concern – all others are non-pathogenic to humans [32]. This chapter describes research performed using a measurement system which is based on low-cost signal acquisition hardware coupled with sophisticated signal processing and classification components. Unlike the previous two chapters (which used e-nose sensors), the system described here is based on Raman spectroscopy, an *optical biosensor* type that was introduced in Chapter 2. This work represents the first time that successful discrimination between *Listeria* species has been performed with Raman spectroscopy.

The work described in this chapter has been disseminated in [23] and [24].

5.1 Background

Contamination of food with *Listeria monocytogenes* continues to be a major problem for consumers and food safety authorities. Unlike most other bacteria, *L. monocytogenes* can grow at refrigerated temperatures and high salinity levels (and can therefore survive and even thrive in conditions generally used to arrest bacteria growth in prepared foods) [104]. *Listeriosis* is the disease that results from infection due to *L. monocytogenes*. This type of infection is very rare (resulting in 0.27 infections per 100 000 people) compared

with other foodborne infections (such as *Salmonella*, with 14.92 infections per 100 000 people) [105][106]. This is fortunate, since the consequences of contracting listeriosis can be dire (especially among particularly vulnerable groups such as pregnant women and the elderly). The mortality rate of listeriosis (estimated at 20-30%) is extremely high compared to *Salmonella* (around 1%) [2]. In Canada in 2008, there was a major outbreak of listeriosis due to contamination in a meat processing facility which resulted in the deaths of twenty people [107]. There are five other species of *Listeria* that occur naturally and pose minimal risk to human health – *L. grayi*, *L. innocua*, *L. ivanovii*, *L. seeligeri* and *L. welshimeri*.

The current gold standard method for detection of *L. monocytogenes* follows closely the conventional methods outlined in Sections 2.1 and 2.2 (culturing and enrichment followed by a battery of biochemical tests – hemolysis, motility, carbohydrate utilization, Gram staining, etc.). This being the case, it suffers from the shortcomings already described, particularly the time required to confirm a positive result. Therefore, developing methods of discriminating more rapidly between *Listeria* species (and especially identifying *L. monocytogenes*) is of real interest to food safety authorities [39][34].

In Section 2.5, an overview of Raman spectroscopy was provided, and its use in the area of bacteria detection and identification was reviewed. The interest in this field has been greatly augmented due to the introduction of a variation on the technique, called *surface-enhanced Raman spectroscopy* (or SERS) [108]. As mentioned already, spontaneous Raman photon emissions are very weak, resulting in a poor signal-to-noise ratio in the measured spectrum. With SERS, the sample under test (e.g., bacteria in this

case) is combined with a colloidal mixture³ containing nanoscale particles (usually made from silver), and the molecules in the sample adsorb onto the surface of these particles, as shown in Figure 5.1. The consequence is that the Raman effect is amplified by several orders of magnitude when irradiated with incoming laser light.

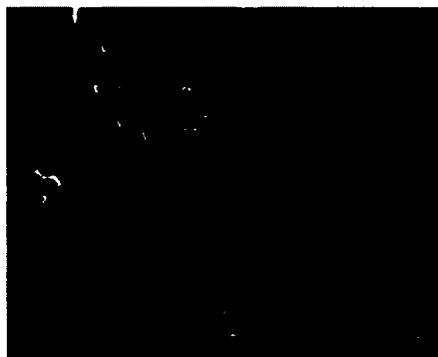


Figure 5.1: Scanning electron microscope image showing the manner in which bacteria (larger) adsorb onto the surface of nanoscale particles (smaller) in a silver colloid mixture in a SERS measurement technique. Reprinted with permission from:

Roger M. Jarvis and Royston Goodacre, "Discrimination of Bacteria Using Surface-Enhanced Raman Spectroscopy", *Anal. Chem.* 2004, 76, 40-47. Copyright 2004 American Chemical Society.

Though the work reviewed has shown that distinguishing between bacteria families is readily possible with SERS, inter-species discrimination (i.e., within the same genus) is more difficult, as shown in Figure 5.2. Because the *Listeria* species are very closely related at the molecular level, their spectral peaks have predominantly the same locations. However, the literature suggests that species-level discrimination of *Listeria* with SERS is plausible. For example, in [109] various *strains* (strain being the taxonomic level below species) of *E. coli* were differentiated with a benchtop Raman instrument and clustering analysis (specifically, a composite dendrogram based on HCA).

³ A **colloid** is a substance microscopically dispersed evenly throughout another substance.

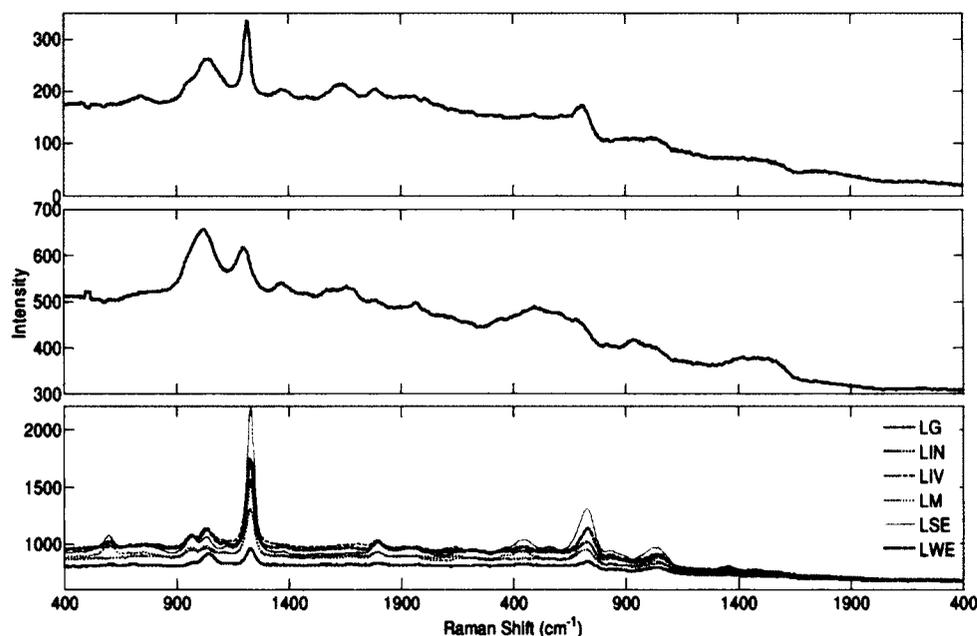


Figure 5.2: Typical SERS spectra for *Salmonella* (top), *Escherichia* (middle), and six species of *Listeria* (bottom).

The experimental work in this chapter commences with Section 5.2, which provides a detailed description of the methods used for generating, processing and classifying the Raman spectra. This is followed by Section 5.3, which demonstrates the results of this work, and Section 5.4 presents an overall summary and conclusions.

5.2 Methods

5.2.1 Sample Preparation

The six *Listeria* species were cultured in BHI broth at 37°C for 16-18h. Bacterial cells were washed three times and then re-suspended in PBS to establish a concentration of 1.67×10^{10} cells/mL. Following this, various spots containing 6 μ L of bacteria suspension and 6 μ L of the silver colloid were deposited on a CaF₂ optical window, mixed in-place with pipette aspiration and ejection, and air-dried for two hours. Additional details of

silver colloid preparation are provided in [110]. Table 5.1 shows the number of SERS spectra from each species used, which included a larger number of *L. monocytogenes* (in order to better characterize the pathogenic species). Bacteria sample preparation was performed by personnel at CFIA.

LISTERIA SPECIES	NUMBER OF SERS SPECTRA, N_s
<i>L. monocytogenes</i> (LM)	120
<i>L. grayi</i> (LG)	71
<i>L. innocua</i> (LIN)	72
<i>L. ivanovii</i> (LIV)	72
<i>L. seeligeri</i> (LSE)	74
<i>L. welshimeri</i> (LWE)	74
TOTAL (N)	483

Table 5.1: Summary of sample numbers and category labels used for experiment in this chapter.

5.2.2 Raman Spectra Measurement

The hardware used to measure Raman spectra in this work was developed by CFIA researchers and described in [110]. Figure 5.3 illustrates the entire measurement system (including hardware and software components). The Raman spectral measurement apparatus included: a) a semiconductor laser source (peak wavelength 784.8nm, operating at 75mW) to illuminate the bacteria sample, b) a fiber optic probe to focus the incident laser light on the sample and to collect the scattered light, c) a mini-spectrometer with reflective grating, CCD to disperse/capture the Raman spectrum coupled with an A/D converter, and d) a PC to store the digitized spectra. It was found experimentally that focusing the laser light on the *edge* of the dried sample spots led to maximum measured intensity, so this was the method used. Raman spectral measurements were performed by personnel at CFIA.

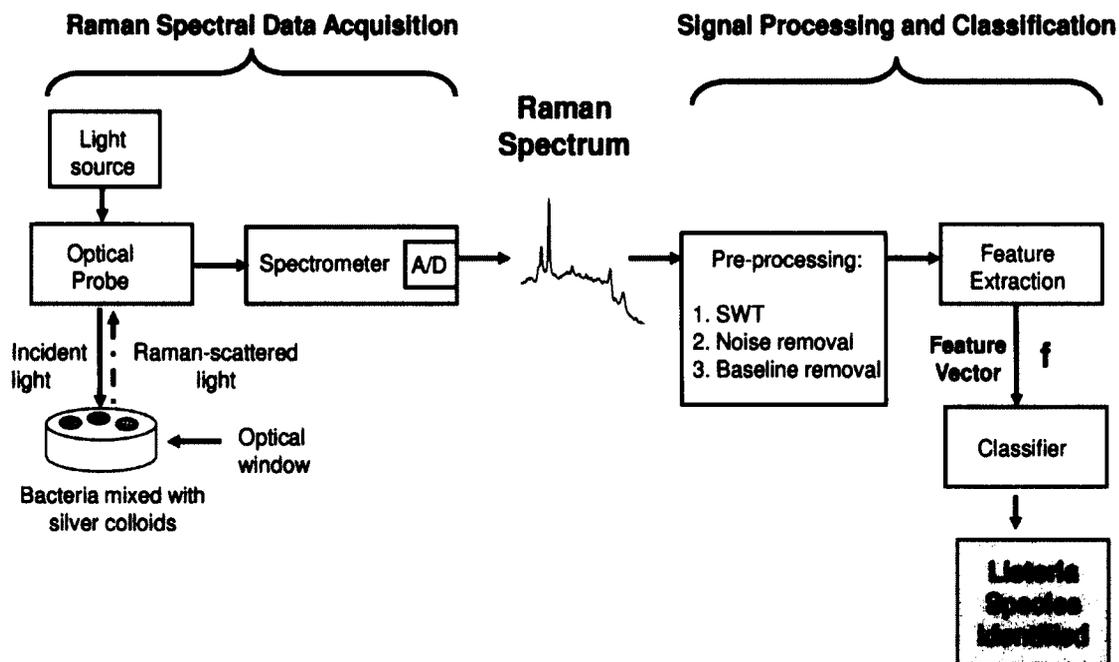


Figure 5.3: Block diagram of overall Raman measurement system to identify *Listeria* species. Shown are both hardware components (left) and software components (right).

5.2.3 Raman Spectral Pre-processing

A wavelet-based transformation was employed for Raman spectral processing because it is effective in separating significant signal features from interfering components in non-stationary signals with high-frequency content (such as the peaks in a SERS spectra) [111][112]. The wavelet representation is described below before proceeding to details about the measured spectra.

A wavelet transform provides an alternative depiction of an input signal $g(x)$ at a given decomposition level L . The wavelet representation consists of [113]:

- a single vector (subband) of M approximation coefficients, denoted $\mathbf{a}_L = [a_1 \ a_2 \ \dots \ a_M]$, representing a low resolution version of the input signal, and
- a set of L vectors (subbands), each containing M detail coefficients, denoted

$\mathbf{d}_l = [d_{1,l} \ d_{2,l} \ \dots \ d_{M,l}]$ for $l = 1, \dots, L$. These represent spectral details at various scales.

From the wide array of wavelet transforms that exist, the stationary wavelet transform (SWT) was selected [114]. The SWT gives the same number of samples in each coefficient subband as appears in the original spectrum ($M = 1600$ in this work). In contrast to the discrete wavelet transform (DWT), the SWT provides a representation wherein the spatial location of information in the Raman spectrum corresponds to the same coefficient location in each of the subbands. The SWT operation can be represented by:

$$\begin{aligned} SWT_L(g(x)) &= [\mathbf{a}_L \ \mathbf{d}_L \ \mathbf{d}_{L-1} \ \dots \ \mathbf{d}_1] & (5.1) \\ &= [a_1 \ \dots \ a_M \ d_{1,L} \ \dots \ d_{M,L} \ d_{1,L-1} \ \dots \ d_{M,L-1} \ \dots \ d_{1,1} \ \dots \ d_M] \end{aligned}$$

where SWT_L is the SWT operator at a given level of decomposition, L . The detail coefficients measure the correlation between the Raman spectrum and a collection of compactly-supported oscillatory basis functions (called *wavelet functions*) differing through the operations of scaling and translation. The coefficient subbands are computed efficiently with a finite impulse response (FIR) filter bank structure, the filter coefficients being derived from the chosen wavelet basis, which in this work was a six-level decomposition ($L = 6$) with a Symlet basis of order 8 (previously used for analysis of Raman spectra [115]). Periodic extension was used at the spectrum's boundaries to calculate wavelet coefficients. The wavelet representation (showing both approximation and detail coefficients) for an arbitrary L *grayi* spectrum is shown in Figure 5.4.

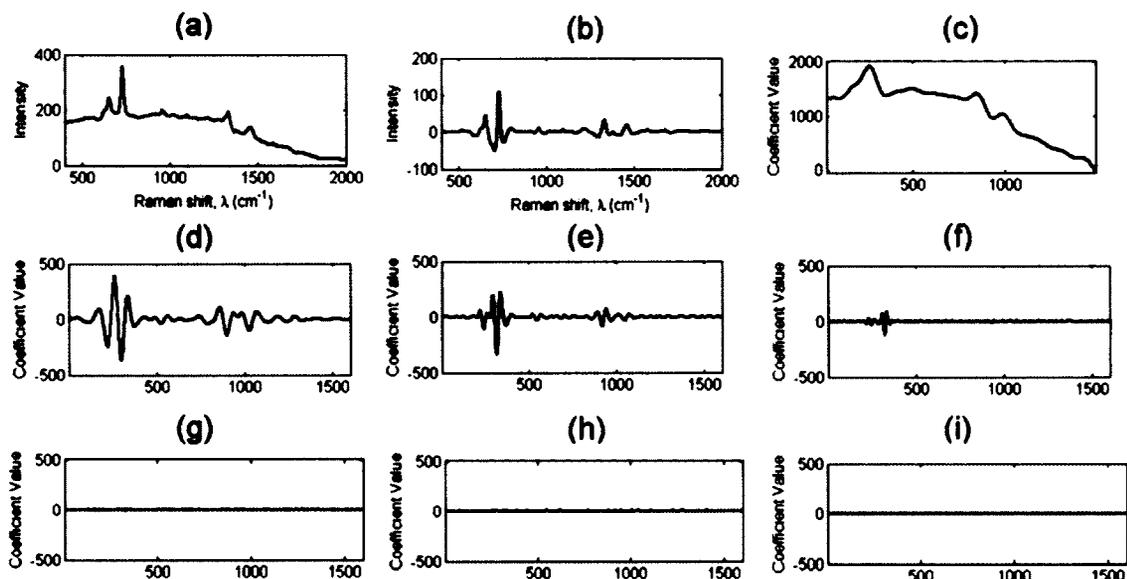


Figure 5.4: Wavelet transform coefficients at different levels of decomposition for an arbitrary *L. grayi* spectrum: a) raw SERS spectrum (after truncation and interpolation), b) preprocessed SERS spectrum, $R_{pre}(\Delta\omega)$ c) a_6 coefficients, d) d_6 coefficients, e) d_5 coefficients, f) d_4 coefficients, g) d_3 coefficients, h) d_2 coefficients, i) d_1 coefficients. The detail coefficients at levels 1-3 ((g)-(i)) have insignificant amplitude and were not used in subsequent processing.

The raw measurement for each sample was an irregularly sampled spectrum of SERS intensity vs. Raman shift ($\Delta\omega$) measured between $130 \leq \Delta\omega \leq 2380 \text{ cm}^{-1}$. After truncation and interpolation (using cubic splines), a new signal $R(\Delta\omega)$ was formed from $401 \leq \Delta\omega \leq 2000 \text{ cm}^{-1}$ with uniform spacing of 1 cm^{-1} (leading to 1600 points per spectrum). By employing the SWT, two dominant interfering components present in $R(\Delta\omega)$ could be removed (see Figure 5.4):

1. High-frequency noise was removed using wavelet-based denoising (which thresholds the wavelet coefficients) using Donoho's universal threshold [116], and
2. Baseline variation (due to the aforementioned fluorescence process) was removed by setting the level 6 approximation coefficients (a_6) to zero, and then reconstructing the signal with the inverse SWT, resulting in the preprocessed spectrum $R_{pre}(\Delta\omega)$ for subsequent processing. In contrast to other background subtraction methods (such as

polynomial fitting [76]), this method does not require specification of any free parameters. No further normalization was performed.

5.2.4 Feature Vector Extraction and Dimensionality Reduction

Below is a description of how $R_{pre}(\Delta\omega)$, measured for each *Listeria* sample, was represented with feature vector \mathbf{f} for later classification. After the SWT was applied to $R_{pre}(\Delta\omega)$, the \mathbf{d}_6 , \mathbf{d}_5 and \mathbf{d}_4 subbands were retained to form the detail coefficient vector for each sample:

$$\mathbf{d} = [\mathbf{d}_6 \quad \mathbf{d}_5 \quad \mathbf{d}_4] = [d_{1,6} \cdots d_{1600,6} \quad d_{1,5} \cdots d_{1600,5} \quad d_{1,4} \cdots d_{1600,4}] \quad (5.2)$$

The feature vector for each sample consisted of a subset of these detail coefficients – those that coincided with the location of peaks in the wavelet domain representation (among all categories) as described below.

5.2.4.1 Determination of Peak Locations for Feature Vectors

This was a training step, performed only once in order to determine relevant wavelet-domain peak locations based on the measured data. The locations of these wavelet peaks provide interspecies discriminatory information that was exploited in downstream classification.

To start, the mean detail coefficient vectors (\mathbf{d}_{LM} , \mathbf{d}_{LG} , \mathbf{d}_{LIN} , \mathbf{d}_{LIV} , \mathbf{d}_{LSE} , and \mathbf{d}_{LWE}) were formed by averaging the detail coefficient vectors for all spectra belonging to that species. For example:

$$\underline{\mathbf{d}}_{LM} = \left| \frac{1}{N_{LM}} \left[\sum_{i=1}^{N_{LM}} \mathbf{d}_6 \quad \sum_{i=1}^{N_{LM}} \mathbf{d}_5 \quad \sum_{i=1}^{N_{LM}} \mathbf{d}_4 \right] \right| \quad (5.3)$$

The summation above is done pointwise over all $1600 \times 3 = 4800$ coefficients considered. From each of these mean detail coefficient vectors (see Figure 5.5), the locations of the peaks in each were determined with a simple local maximum search. These peak locations were denoted by the sets E_{LM} , E_{LG} , E_{LIN} , E_{LIV} , E_{LSE} , and E_{LWE} . In order to assemble a candidate set of peak locations, the union of these, E_{all} , was computed. Taking the union was done to extract discriminatory information (since spectra from one species may exhibit peaks at locations not present in those of the others). The next two stages removed certain peak locations from E_{all} :

1. If a peak location in E_{all} had very low amplitude coefficients *over all six categories*, it was removed since it would be of little value in differentiating species. An empirically determined threshold of $T_A = 25$ (See Figure 5.5) was used to reject these peak locations.
2. For a given range of detail coefficients, the mean vectors \mathbf{d}_{LM} , \mathbf{d}_{LG} , etc. exhibited slight variations in the location of their peaks (See Figure 5.5). To ensure that only one location was retained in this range, a simple clustering was performed based on thresholding the differences between peak locations. This eliminated several peak locations in E_{all} which were in very close proximity but still represented the same overall maximum.

E_{final} was formed from E_{all} by removing the locations identified in the two steps above. E_{final} represents the location of peaks that occur in the wavelet representation of at least one species.

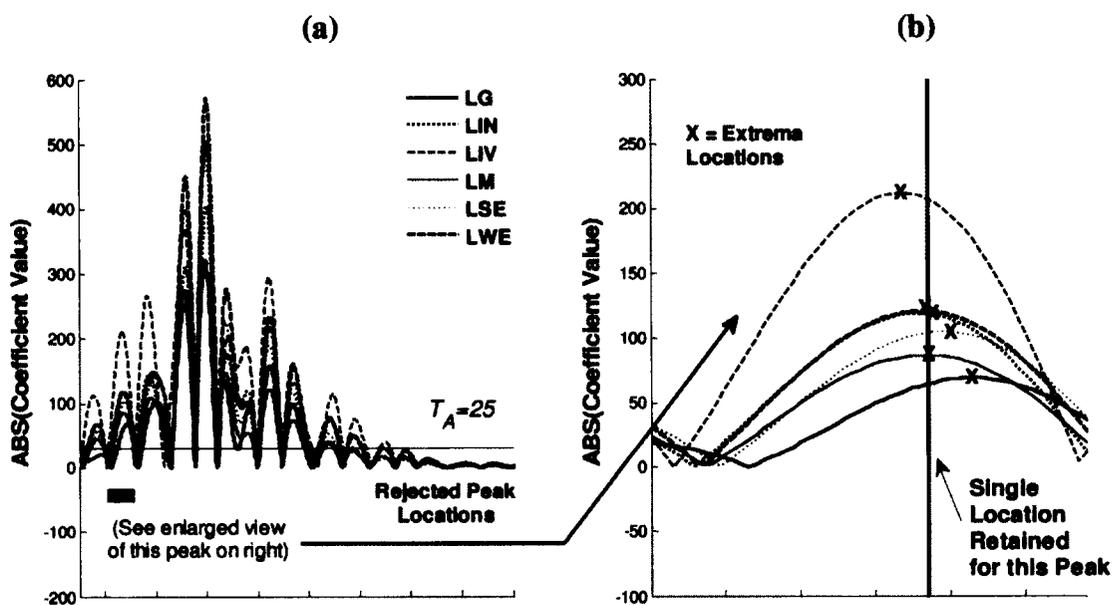


Figure 5.5: Feature Extraction from SERS spectra. All samples in each of the six categories were used to generate the detail coefficient averages. After finding the location of all local maxima (for all categories), peak locations were eliminated through (a) amplitude thresholding and (b) clustering. The peaks in (b) correspond to an enlarged view of the underlined peaks in (a). For clarity, only a subset of \mathbf{d}_6 is shown.

5.2.4.2 Selection of SWT Coefficients at Peak Locations

To create the feature vector \mathbf{f} for an unseen bacteria sample's spectrum, the amplitude of the detail coefficients at the peak locations in E_{final} were calculated as follows:

1. The raw spectrum was preprocessed, yielding $R_{pre}(\Delta\omega)$ as described in Section 5.2.3,
2. The SWT of $R_{pre}(\Delta\omega)$ was calculated and only \mathbf{d}_6 , \mathbf{d}_5 and \mathbf{d}_4 subbands were retained, yielding \mathbf{d} as described in Section 5.2.4, and
3. The values of \mathbf{d} at the coefficient locations in E_{final} were extracted, yielding \mathbf{f} .

The above process generated feature vectors of dimension 74 (for six-category classification) and 55 (for two-category classification, *L. monocytogenes* vs. all other species). Following this, dimensionality reduction using PCA and MDA (described in Chapter 3) was applied.

5.2.5 Classification and Performance Characterization

As with previous work described in this thesis, an LDA classifier was used. With linear decision boundaries, this classifier is less likely to be overtrained (which occurs when the system performs well on training samples, but does not generalize well for unseen test samples). The classification performance was validated with:

1. **LOO cross-validation** (See Section 3.2.5), and
2. **Bootstrap Resampling**. The relatively high variance inherent to the LOO method can be decreased by calculating a *bootstrap estimate* of classification accuracy [90][117]. With this method, a set of $N_{train,boot}$ samples (the training set) are chosen randomly *with replacement* from the complete set of N samples (from all categories) with equal probability (uniform distribution). Because the training samples are drawn with replacement, duplicate samples will almost certainly exist in the training set. The remaining $N_{test,boot}$ samples (the test set) are used to test the classifier. This process is repeated (with different test and training sets) for a high number of bootstrap replicates, N_B , and at each iteration i , the per-iteration classification accuracy is calculated, CA_i . In this work, $N_B = 500$ was used, which represented a compromise between reduced CA variance estimate and computational efficiency. The mean (CA_{Boot}) and standard deviation (SD_{Boot}) are computed as follows:

$$CA_{Boot} = \frac{1}{N_B} \sum_{i=1}^{N_B} CA_i \quad SD_{Boot} = \sqrt{\frac{1}{(N_B - 1)} \sum_{i=1}^{N_B} (CA_i - CA_{Boot})^2} \quad (5.4)$$

5.2.6 Feature Ranking

The extracted features are spatially aligned with those in the original Raman spectrum (this is a property of the SWT used). This property was used to identify the features that perform best when distinguishing *Listeria* species. For multi-category data set ($C = 6$ species), there are two feature ranking options:

- 1) consider all possible category pairs ($6!/((6-2)!(2!)) = 15$ in this case), then find the best features to distinguish each pair, or
- 2) consider six separate binary classifications, one for each species. For each classification, there are two groups: a) the spectra from a single *Listeria* species and b) all others.

In this work, the latter approach is used. Note that this method of ranking features is independent of any classifier. Ranking features best able to distinguish species c from the others is done by calculating the following statistic for each feature f :

$$z_f = \frac{|m_c - m_{\bar{c}}|}{\sqrt{\frac{s_c^2}{N_c} + \frac{s_{\bar{c}}^2}{N_{\bar{c}}}}} \quad (5.5)$$

where m_c and s_c are the mean and standard deviation of the N_c samples in category c , and $m_{\bar{c}}$ and $s_{\bar{c}}$ are the mean and standard deviation of the other $N_{\bar{c}}$ samples. Higher z_f values indicate good discrimination from those features, since they have a larger separation of the means (relative to the variances) [118].

5.2.7 Confidence Measures

The confidence measures described in the previous chapter were also applied to this classification problem (see Section 4.2.5).

5.3 Results

The overall system accuracy is shown in Table 5.2 (six-category problem, CA = 96.1% using all five MDA components as inputs to the classifier) and Table 5.3 (two-category problem, CA = 97.9% using all 55 features as inputs to the classifier), along with details on misclassifications. These results represent a sensitivity of 95% and a selectivity of 98.9% for discrimination of *L. monocytogenes* from all other species. Tables 5.2 and 5.3 demonstrate that the system's classification performance was substantially better than what would be expected from random guessing – 16.7% and 50%, respectively.

		CLASSIFICATION RESULT						CA (%)
		LG	LIN	LIV	LM	LSE	LWE	
TRUE CLASS	LG	70	1					98.6
	LIN	1	69		1		1	95.8
	LIV	1	1	70				97.2
	LM		2		113	3	2	94.2
	LSE	1				72	1	97.3
	LWE		1		1	2	70	94.6
OVERALL CA								96.1

Table 5.2: Classification results and confusion matrix for identification of six species of *Listeria*. Per-category classification accuracies appear in the rightmost column. Entries in the shaded area represent correct classifications; misclassifications appear elsewhere. Blanks entries indicate zero and $T_A = 25$ was used. Classification was performed using all 5 MDA components.

		CLASSIFICATION RESULT	
		LM	Other
TRUE CLASS	LM	114	6
	Other	4	359
OVERALL CA:		97.9 %	

Table 5.3: Classification results for identification of *L. monocytogenes* amongst all species, $T_A = 25$. Classification was performed using all 55 input features.

The feature ranking results are shown in Table 5.4. The most distinguishing features arise from the smaller-valued detail coefficients in the d_5 and d_4 subbands. This makes intuitive sense – the coarse scale features represented in the d_6 subband are not alone capable of discriminating species; the additional information in d_5 and d_4 are required to complete the picture. This provides evidentiary support for the use of wavelets (a form of multi-scale analysis) for characterizing *Listeria* spectra.

SPECIES		MOST DISCRIMINATING FEATURES (DECREASING VALUES OF τ_l)
Six- category	LG	75 (D6), 205 (D5), 676 (D6), 637 (D6), 630 (D6), 64 (D5), 86 (D5), 42 (D5), 687 (D6), 72 (D4)
	LIN	431 (D6), 292 (D4), 807 (D6), 114 (D6), 374 (D6), 215 (D4), 311 (D4), 274 (D5), 205 (D5), 367 (D4)
	LIV	679 (D5), 630 (D6), 223 (D5), 244 (D5), 676 (D6), 521 (D5), 42 (D5), 75 (D6), 205 (D5), 221 (D6)
	LM	244 (D5), 223 (D5), 215 (D4), 292 (D4), 523 (D6), 245 (D4), 560 (D6), 232 (D4), 1042 (D5), 274 (D5)
	LSE	274 (D5), 205 (D5), 215 (D4), 292 (D4), 186 (D5), 311 (D4), 916 (D5), 295 (D5), 927 (D4), 807 (D6)
	LWE	679 (D5), 803 (D5), 72 (D4), 1168 (D5), 983 (D6), 86 (D5), 64 (D5), 1190 (D6), 42 (D5), 274 (D5)
Two- Category	LM vs. Others	243 (D4), 244 (D5), 230 (D4), 291 (D4), 223 (D5), 215 (D4), 521 (D6), 558 (D6), 1041 (D5), 914 (D5)

Table 5.4: Feature Ranking (no DR applied). Each row shows the 10 best features used to distinguish *Listeria* species from all others. The number refers to the Raman shift position (cm^{-1}) in the original spectrum from which the feature was derived, and the (Dl) in parentheses indicates that the feature resulted from the coefficients at the l^{th} decomposition level. Six-category and two-category cases are shown.

Figure 5.6 shows the effect on the accuracy of the *Listeria* classification system when thresholds on the posterior- and silhouette coefficient-based confidence measures are applied. As was seen in Chapter 4, the CA value increases with the rejection ratio. The ARC curves (also shown) could be used to determine the operating point (i.e. value of the threshold) that represents the optimal tradeoff.

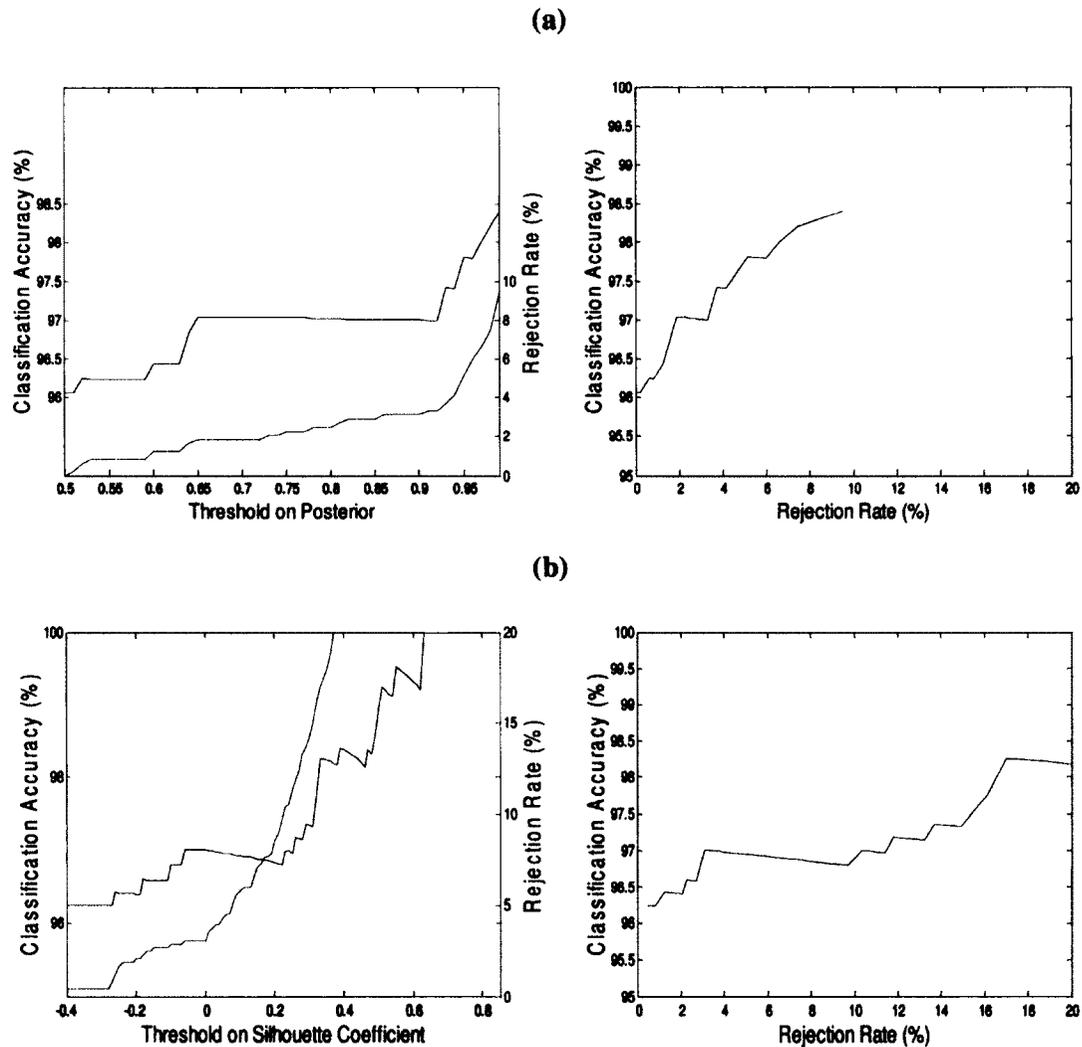


Figure 5.6: Effect of varying confidence threshold on classification accuracy and rejection rate: (a) posterior probabilities and (b) silhouette coefficients. In the ARC on the right, each data point corresponds to a unique threshold value.

Figure 5.7 illustrates the effect of applying DR to the feature vectors on the system's classification accuracy (using bootstrap estimates). Without DR (using the raw feature vectors), a high number of components are needed for reasonably high classification performance (greater than 40 components for $CA \geq 90\%$, an arbitrarily chosen value). The unsupervised DR method (PCA) outperformed the no DR case (for a fixed number of inputs), in some instances by upwards of 20% improvement. This demonstrates that the original feature vectors contain redundant information that is suppressed with PCA, which reduces the correlation between components. Superior performance is seen with the supervised DR method (MDA), which incorporates the samples' category information in the resulting transformation. With MDA, the feature vector dimension can be drastically reduced (to only four components) in order to achieve the same arbitrary $CA \geq 90\%$ target. For all three methods, as the number of components increases, the standard deviation of the CA estimates decreases (showing that the bootstrap procedure results in a more accurate CA estimate).

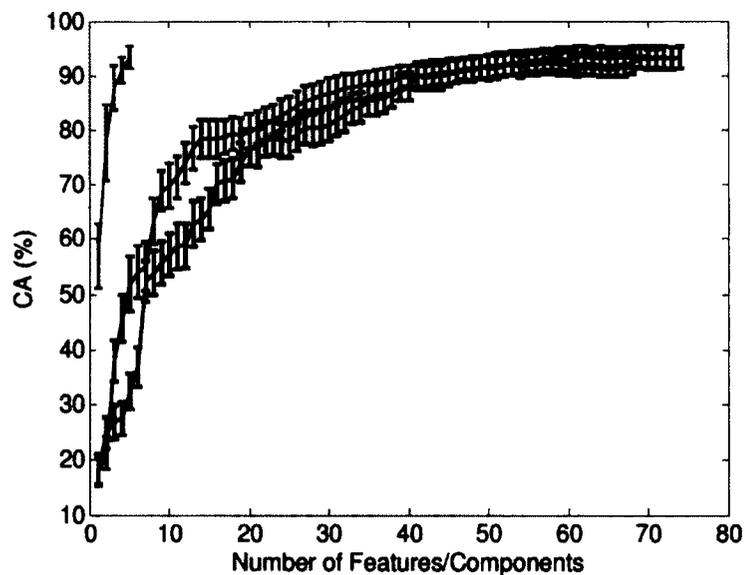


Figure 5.7: Effect of DR on classification accuracy (with bootstrap resampling), six-category case, $T_A = 25$, $N_{train, boot} = N = 473$. Error bars show $CA_{Boot} \pm SD_{Boot}$. With MDA, only $C - 1 = 5$ features are possible. DR using MDA is shown in red; using PCA is shown in blue; no DR is shown in black.

Figure 5.8 allows visualization of the effect of applying DR to the feature vectors. In these graphs, the high dimensional feature vector is reduced to only two components. The supervised MDA technique provides superior separation between categories over PCA and helps to illustrate why MDA give better classification accuracy results (for a given number of components/features). Note that the highest classification rates were achieved with a larger number of components than shown in this figure (thus very difficult to visualize).

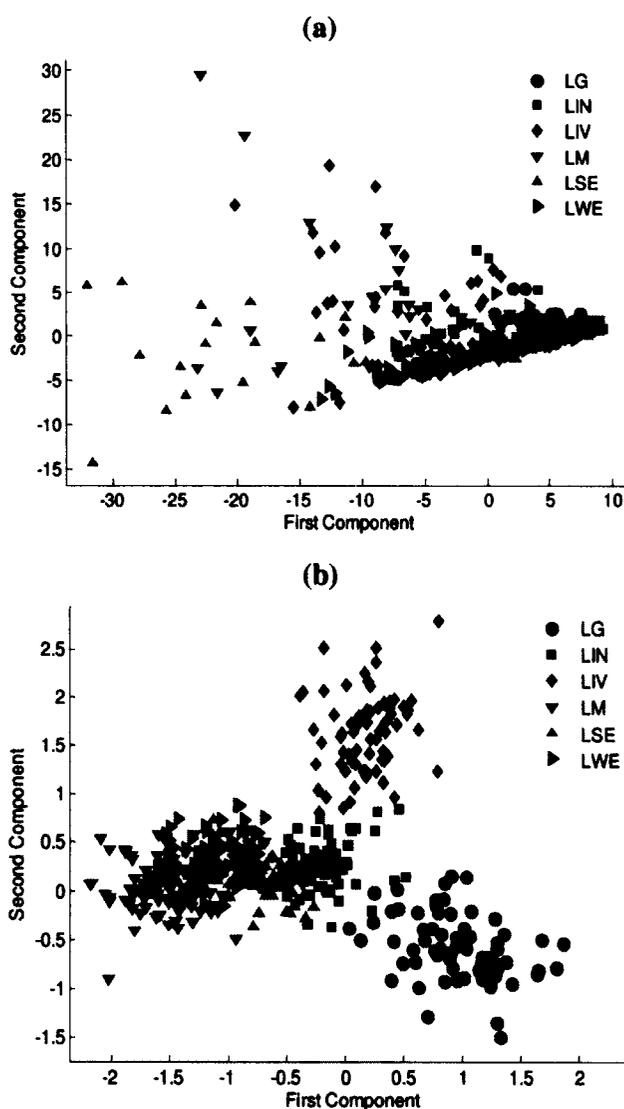


Figure 5.8: Clustering in DR space (six-category case): a) PCA b) MDA. Each sample point represents a *Listeria* SERS spectrum after DR has been applied. In each case, two components are retained. $T_A = 25$ was used.

Recall that in the method described, the threshold $T_A = 25$ was used to filter out features with insignificant coefficient amplitude. Decreasing this value will result in a higher-dimensional feature vector; conversely, increasing the threshold will give a feature vector of lower dimension. Figure 5.9 demonstrates, though, that CA remains high over a wide range of T_A and the corresponding number of retained features ($CA \geq 90\%$ for $5 < T_A < 100$). This illustrates a very desirable property of the system – the performance is not overly sensitive to the free parameter selection. It can also be seen from this figure that the benefit of using MDA to reduce the dimensionality of the feature vector also applies across this wide range.

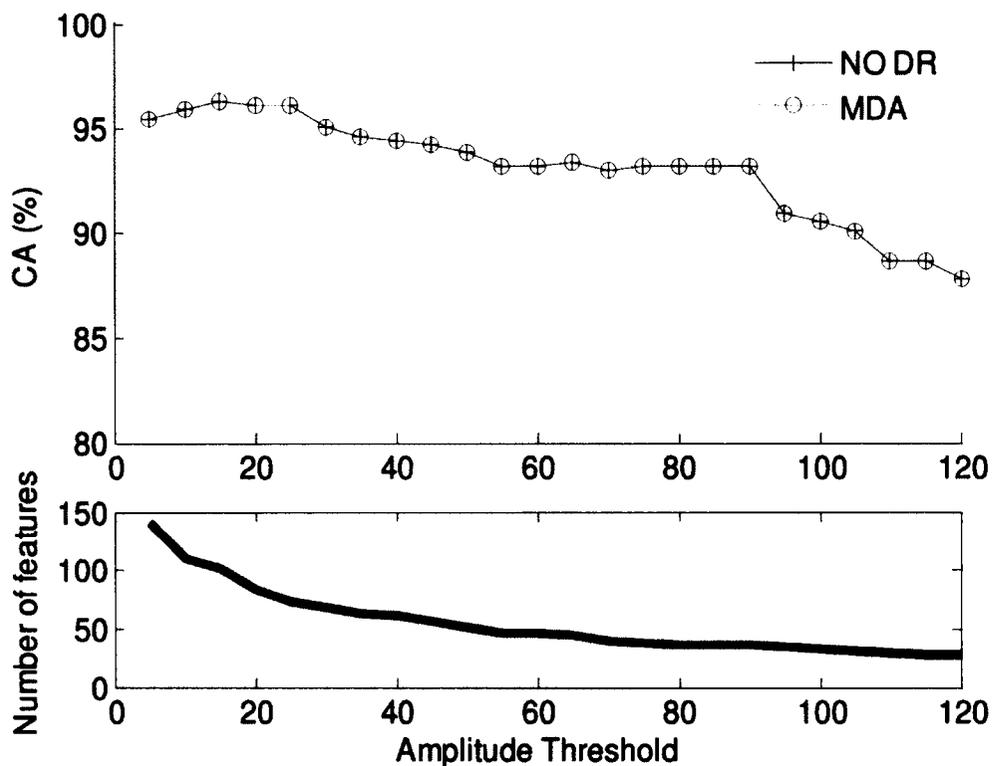


Figure 5.9: Effect of varying T_A on i) CA with DR (MDA, using 5 components) – top plot, and ii) number of retained features – bottom plot. Results are for the six-category case using LOO CV.

5.4 Discussion and Conclusion

In this chapter, work was described that demonstrated (for the first time) that SERS technology is capable of differentiating between *Listeria* species. The development of a SERS measurement system was described, which used portable, low-cost Raman spectral acquisition hardware, coupled with signal processing and classification software. This system could readily be packaged for field or on-site food safety applications. The proposed analysis method uses a wavelet-based representation of the SERS spectrum to a) remove noise and background artifact and b) extract discriminating features based on peak locations in the wavelet domain. The classification system used a linear classifier in conjunction with dimensionality reduction and bootstrap resampling. Accuracies of over 90% were reported for both 6-category (identification of *Listeria* species) and 2-category (identification of *L. monocytogenes*) problems.

The system proved to be robust to variations in measurement conditions. In preliminary work on the same application, 167 *Listeria* spectra were collected and analyzed [24]. Samples used in the present work included these spectra, with the remaining $(N - 167) = 316$ spectra collected under different measurement conditions (e.g., different batch of silver colloids, different bacterial cultures). This increases the confidence in the reliability of the method, particularly in light of the largely increased number of samples in the *L. monocytogenes* category.

The exploratory work described herein motivates further development on a system that might offer several potential advantages over the existing gold standard method for discrimination amongst *Listeria* species. The bacteria samples employed in this work are

available relatively early in the process, as shown in Figure 5.10. Once the initial colonies have been isolated (the first stage, which must still be performed), identification of the *Listeria* species can be obtained within 24h. This rapid result is obviously important to food safety officials, but also offers several operational advantages such as reduced cost for reagents, media, and labour. It is hoped that this research demonstrates the potential for this rapid fingerprinting technique to become a technology that complements existing biochemical gold standards (in a confirmative or pre-screening role). In doing so, it could help eventually help circumvent a principal problem with conventional methods described in Chapter 2.

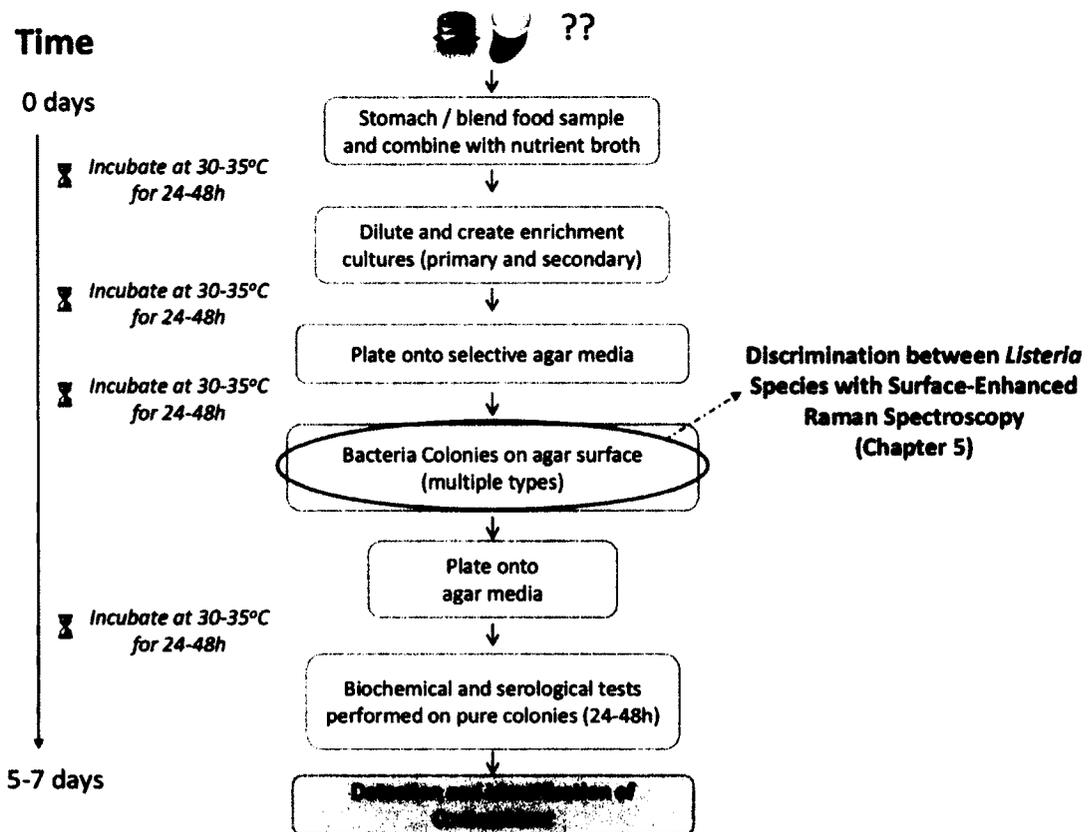


Figure 5.10: Simplified view of conventional methods. Highlighted block indicates the stage at which the sample type considered in Chapter 5 is available.

Chapter Six

Electronic Nose-based Tracking of Spoilage in Real Food Samples (with possible applications to Smart Home)

The measurement systems described in the last several chapters used input samples based on *pure bacteria colonies* (after an incubation phase), either suspended in nutrient broth (Chapter 3), suspended in PBS (Chapter 4), or mixed with colloidal silver (Chapter 5). In contrast, this chapter describes experiments that were performed using *unprocessed food* as the input sample, the overall goal being the ability to identify foods that are unsafe to eat based on odour (in a *smart home* setting). The food samples were left to spoil for several days (under various conditions), and the evolving odour signatures were measured with an e-nose. It should be acknowledged at the outset that the presence of bacteria (the focus of previous chapters) represents *only one of many simultaneous mechanisms* that might affect the volatile patterns resulting from spoiling food. Other phenomena that might contribute are the development of yeasts, moulds, etc. Since this work also represents the first time that an exploratory analysis into the feasibility of using e-nose technology for smart home applications (as a potential assistive technology for older adults) has been performed, a technical framework for such an application is introduced at the end of the chapter.

The work described in this chapter has been disseminated in [25–27] and the following journal manuscript has been submitted:

G. C. Green, A. Chan, and R. A. Goubran, “Integration of Odour Monitoring into a Smart Home”, submitted to *IEEE Transactions on Information Technology in Biomedicine*.

6.1 Background

The ability to determine the safety of a food product at its “point of use” is highly desirable. Unfortunately, detecting the presence and nature of contaminants in unprocessed food is extremely difficult (described by one organization as a “needle in a haystack” problem [119]). This is due to the highly complex nature of the food matrix and the small number of pathogenic organisms that can render the product unsafe. There has been some progress in this regard, notably the introduction of the so-called real-time PCR (e.g., TaqMan probes by Applied Biosystems). Despite its name, this method still requires upwards of 24h to return an unequivocal result [120].

If the future development of a real-time *in situ* food spoilage detector proves to be technically feasible⁴, one demographic that would stand to benefit significantly would be *older adults*. As mentioned in Chapter 1, this group may be less able to identify spoiled foods (e.g., due to dementia or impaired sense of smell) and this is of particular concern to this age group, since weakened immune systems make them more susceptible to contract foodborne illness [6][7].

The work in this chapter was motivated by recent developments in the area of *ambient-assisted living environments*, with a particular interest in the *smart home*. A smart home is defined as:

“a residential setting equipped with a set of advanced electronics, sensors and automated devices specifically designed for care delivery, remote monitoring,

⁴ In 2007, a product called SensorFreshQ (Food Quality Sensor International, Lexington, Massachusetts) was being marketed. SensorFreshQ claimed to be a consumer handheld electronic nose that detected meat and poultry spoilage by directly sniffing the raw product and producing a color-coded indication (red, green, yellow) as to whether or not it was safe to be consumed. (It appears that SensorFreshQ is no longer available for purchase).

early detection of problems or emergency cases and promotion of residential safety and quality of life.” [121]

In the developed world, older adults represent the fastest growing age group, and this trend is expected to continue [122]. In Canada, for instance, the percentage of the population over the age of 65 is expected to grow from 14% to 26% by 2050 [123]. In most instances, these people would prefer to live in their homes as they age (known as *aging in place*). This has advantages both for the older adult (e.g., more independence, ability to maintain social life), their families (e.g., saving the cost of a nursing home, which is a common alternative) and society in general (e.g., avoiding increased cost of long-term hospital care on already strained health care budgets). Unfortunately, older adults are particularly vulnerable to certain health problems (e.g., frailty, dementia, or mobility impairments) that make it difficult to continue to live at home safely.

There has been increased interest in using technology to assist with this goal. In particular, researchers are introducing novel sensor modalities and assistive devices into older adults’ residences to help them adapt to these new challenges while continuing to live at home. Certain devices such as wearable vital sign monitors and assistive lighting are already being used. Other candidate technologies are being investigated for this type of use in projects such as TAFETA Smart Systems for Health, a research program at Carleton University and the Elizabeth Bruyère Research Institute in Ottawa, Canada. TAFETA (Technology-Assisted Friendly Environment for the Third Age) brings together multi-disciplinary experts to develop smart technologies that help seniors live independently in safe, responsive environments [124]. In addition to those already mentioned, a variety of other assistive technologies are currently being investigated as

part of the TAFETA project, such as smart grab bars, motion sensors, and microphone arrays [125][126]. Similarly, an e-nose based system that could identify when food is unsafe to eat and notify the resident would be of real benefit, since an older adult not able to tell that a food has spoiled (e.g., due to impaired sense of smell) might ingest it (along with any bacteria responsible for spoilage). Additionally, such a system could recognize whether the presence of spoiled food in the residence is becoming a more frequent or prolonged occurrence (which might warrant the attention of a family member or health care provider).

There is a large literature demonstrating the ability of e-noses to track food spoilage in very controlled environments [127–129]. Generally a single food is considered, and the e-nose sensor outputs are demonstrated to correlate with the mechanism which causes the spoilage (e.g., in [129], the total viable counts of bacteria present in red meat are measured throughout the 15 day spoilage process and the e-nose responses and its classification engine are trained against this gold standard bacteriological method). In contrast, work presented in this chapter focused on a *variety* of common foods (without bacteriological analysis during spoilage). Instead, a very simple measure for spoilage degree was used (essentially, the amount of time during which the food has been left out at room temperature). This decision was deliberate, and allowed the investigation of the suitability of e-nose technology in this application with as simple an arrangement as possible.

This chapter describes a set of three experiments performed to research the applicability of an e-nose based measurement system capable of tracking food spoilage odours in the smart home. Because this was an exploratory work, a number of techniques

were investigated. It was unclear at the outset whether taking consecutive odour measurements during the spoilage period (for a given food) was best done by:

- a) using different aliquots of the spoiling food in different vials (whereby consecutive sniffs were done on unique samples), or
- b) using the same aliquot in a single vial, whereby consecutive sniffs were done on the same sample.

Both had the potential to introduce variability into the measurements. In the former case, the variability would obviously be due to the fact that different aliquots were being sniffed, while in the latter case, the variability would potentially be introduced by the effects of re-processing the same vial repeatedly.

Mimicking the structure of Chapters 3 and 4, the work in this chapter is organized by first describing the methods common to all experiments (Section 6.2). Experiment #1 is then described in full (Section 6.3), in which additional details regarding the background, methods, results and discussion are included. A similar structure follows for Experiment #2 (in Section 6.4) and Experiment #3 (in Section 6.5). In organizing the work in this manner, the reader will realize that the results from each experiment motivated the revisions to the methods used in subsequent experiments. Section 6.6 presents a proposed framework for incorporating e-nose technology in the smart home and includes a detailed discussion of the merits and challenges involved. Section 6.7 presents an overall summary and conclusions of this chapter.

6.2 Methods

The information presented in this section is applicable to all subsequent experiments – these include methods related to e-nose parameters, the use of artificial neural networks, fuzzy clustering and dimensionality reduction.

6.2.1 Electronic Nose Analysis

These experiments used the AlphaMOS FOX 3000 with twelve MOS sensors, as previously described (Chapter 3). A headspace injection volume of 2mL was used in all cases. Sensor responses were recorded every 0.5s for 120s. Headspace generation and injection was performed with the HS100 autosampler. Headspace generation parameters (time and temperature) varied for each of the experiments, as described subsequently.

6.2.2 Data Processing – Artificial Neural Networks

A back-propagation multi-layer perceptron (MLP) – a type of artificial neural network – was used to correlate e-nose feature vectors with the degree of spoilage. In this work, a simple approach was adopted – the training targets (value of the single output node) for each food sample were *linearly related* to the amount of time that it had been left out to spoil. The target for fresh samples (i.e. $t = 0$) was set to zero; for the oldest (i.e. most spoiled) samples, the target was 1.0. The MLP used one hidden layer comprising 10 neurons, and the input layer consisted of 12 neurons (one for each e-nose sensor). The Matlab Neural Network Toolbox numerical computing environment was used (The Mathworks, Natick, MA, USA) [130]. The transfer function “*tansig*” was used for the

hidden layer, and “*purelin*” was used in the output layer. To avoid overtraining, the input patterns were divided randomly into training, validation, and test sets (using 60%, 20%, and 20% of the data, respectively). Network training employed the Levenberg-Marquardt backpropagation algorithm until a mean-squared error target of less than 10^{-6} was achieved, or the validation set error increased [93].

6.2.3 Data Processing – Fuzzy Clustering

As stated above, time (specifically, the number of days at which the food has been left out at room temperature) was used as the measure of the degree of food spoilage. While simplifying the analysis, this assumption does not consider that: a) different foods will spoil at different rates, and b) any numerical measure of spoilage assigned without having performed a bacteriological analysis (e.g., colony counts) of the spoiled foods should be treated with a certain degree of caution. These observations motivated the use of *fuzzy methods*, which assume that the output categories are not mutually exclusive (Day 1, Day 2, etc.). Rather, each sample receives a graded membership in each cluster, typically represented with a scalar u between 0 and 1 [131].

Fuzzy *C*-means (FCM) clustering (which has previously been used with success in e-nose analysis [68]) allows each of the sample patterns to have an association with *multiple clusters* instead of belonging to a single category exclusively. FCM is based on the minimization of the objective function:

$$J_2 = \sum_{i=1}^N \sum_{c=1}^C u_{ic}^2 \|x_i - m_c\|^2 \quad (6.1)$$

where C is the number of categories, x_i ($i = 1, \dots, N$) are the data points, m_c ($c = 1, \dots, C$) are the locations of the cluster centroids, and u_{ic} are the individual elements in the membership matrix U , which represents the levels of “belongingness” of an input data point i to a cluster c [132]. The locations of the cluster centroids and u_{ic} values are updated iteratively according to:

$$m_c = \frac{\sum_{i=1}^N u_{ic}^2 \cdot x_i}{\sum_{i=1}^N u_{ic}^2} \quad u_{ic} = \frac{1}{\sum_{k=1}^C \frac{\|x_i - m_c\|}{\|x_i - m_k\|}} \quad (6.2)$$

until the maximum change in the u_{ic} values at each iteration falls below a pre-defined threshold.

6.2.4 Data Processing – Dimensionality Reduction

Both unsupervised DR (using PCA) and supervised DR (using MDA) were employed, as described previously (see Section 3.2.4).

6.3 Experiment #1 – Feasibility Study with Multiple Foods, Irregular Sampling

6.3.1 Experiment #1: Food Samples, Sample Preparation and Sampling Protocol

The following food samples were acquired from a local grocery store and placed in a 4°C refrigerator prior to the start of the experiments:

- cream (18% MF), category label: C
- eggs, category label: E

- milk (3.25% MF), category label: M
- sour cream (5% fat), category label: S
- yoghurt (6% fat), category label: Y

The foods above were transferred from the original packaging to identical open containers and left at room temperature (roughly 23°C) during the course of the experiment (which lasted 7 days) in order to expedite spoilage. Food samples were processed on five separate days during the experimental period – on days 1 (fresh), 2, 3, 5, and 7. On each of these days, the food was thoroughly mixed, and then four 1mL samples of each food were transferred by pipette into separate 10mL glass vials and immediately capped. These samples were analyzed by the e-nose in alternating order (C,E,M,S,Y,C,E,M,S,Y...). On a given measurement day, consecutive e-nose responses were captured roughly 30min apart. A total of 100 samples were recorded, consisting of 4 vials/category (for each day) x 5 categories x 5 processing days. Figure 6.1 shows the food samples at the beginning and end of the experiments, illustrating food spoilage by the end of the week.



Figure 6.1: Photographs of food samples used in Experiment #1: (a) Day 1 and (b) Day 7.

6.3.2 Experiment #1: Additional Details

During e-nose processing, headspace generation was performed by heating the capped vials containing the food sample for 300s at a temperature of 50°C. Feature vectors were constructed from each MOS sensor with fractional baseline manipulation (see Section 3.2.3.1), extraction of the maximum absolute value of the sensor response curve (see Section 3.2.3.2) and vector normalization to normalize the features for all samples, resulting in vectors in the direction of \mathbf{f}_i with unity amplitude:

$$\hat{\mathbf{f}}_i = \frac{\mathbf{f}_i}{\|\mathbf{f}_i\|} \quad (6.3)$$

6.3.3 Experiment #1: Results

Figure 6.2 shows the PCA representation of each of the food samples on Day 1 and Day 7 (using a PCA transformation matrix calculated with only these 40 samples). On the first day (fresh), the samples are clearly differentiated (indicating unique odours), as would be expected. By the end of the week, however, the spoilage process has had a marked effect on the clustering results. In particular, there is significant overlap between clusters and larger cluster widths. On Day 7 (after a week of spoilage), though milk and eggs remain separated, the odour patterns of other categories (cream, sour cream and yoghurt) exhibit overlap that was not evident at Day 1.

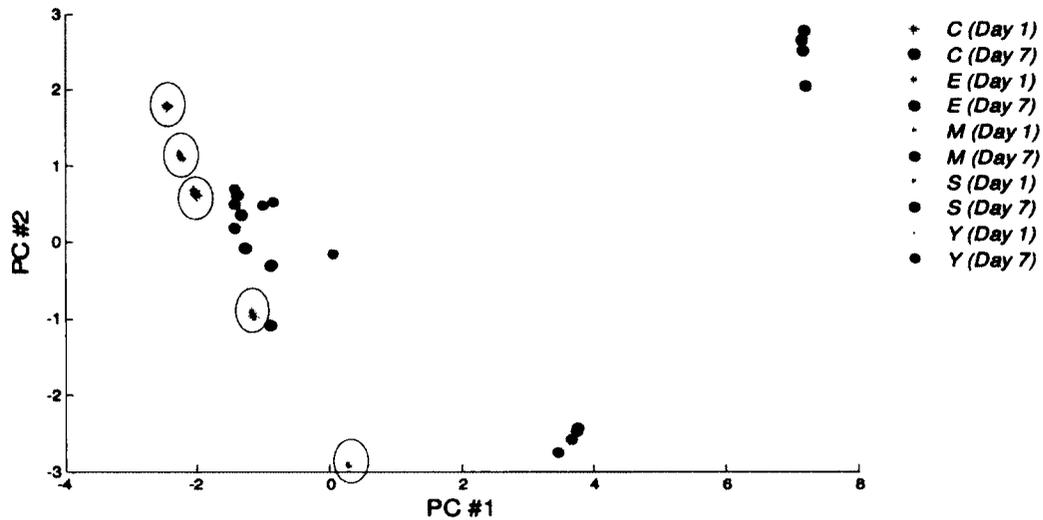


Figure 6.2: Experiment #1: PCA clustering results for all food sample measurements on Day 1 (asterisks) and Day 7 (filled circles). On Day 1, separation between food categories is evident (highlighted with circles). By Day 7, the odour profiles have changed considerably (increased variability and overlap between clusters).

It is instructive to observe the change during the week for each food, as illustrated for a couple of foods in Figure 6.3. (In each of these plots, the PCA transformation matrix was re-calculated using only the 20 samples in that food category). In most cases, the individual clusters (corresponding to the same food sample on different days) show minimal overlap, indicating that there are distinguishable odour patterns associated with the food as they spoil. Additionally, the centroids of these clusters define a relative smooth trajectory (arrows in Figure 6.3). This behaviour is intuitive – the food spoilage process is gradual (caused by, for instance, bacterial growth in this host), so sudden discontinuities or jumps would not be expected [133]. Finally, it was observed that the foods that best exhibited this behavior were eggs, sour cream, and yogurt (the effect was not so evident with milk and cream).

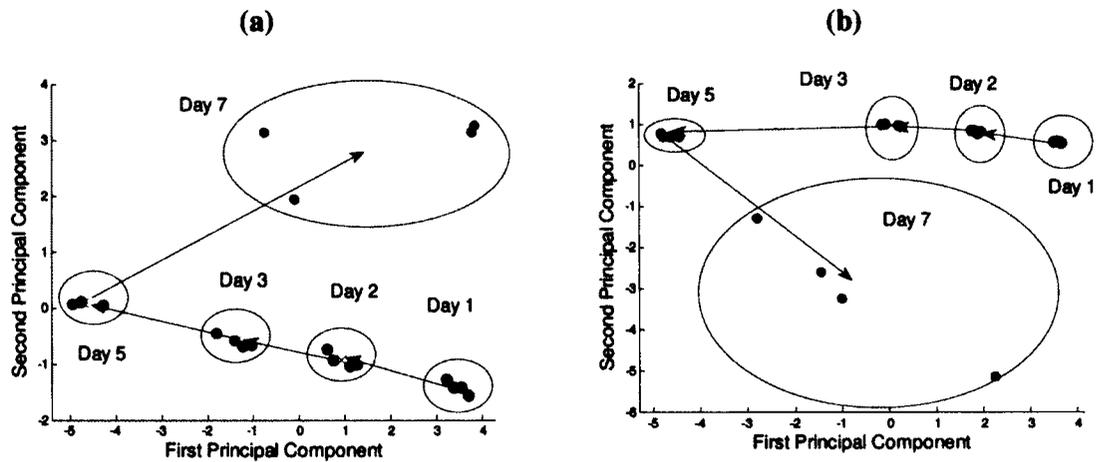


Figure 6.3: Experiment #1: PCA clustering results for food samples measured throughout the entire experiment: (a) Eggs, (b) Sour Cream. Labels indicating the day of measurement are included for each cluster. Cluster centroids are marked with an “X” and the arrows show centroid progression with time.

Figure 6.4 presents a regression analysis showing the MLP outputs along with the specified targets for these input patterns (the data in this plot are shown for yoghurt samples only). The measured correlation coefficient ($R > 0.99$) demonstrates that this particular MLP architecture was able to learn the changing odour patterns as spoilage occurred.

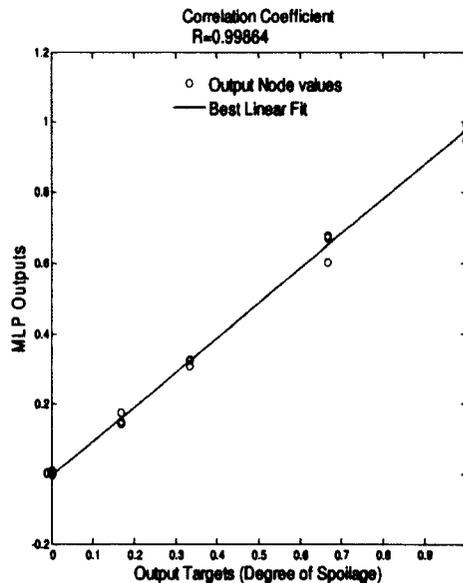


Figure 6.4: Experiment #1 - Results of regression analysis comparing the sample age - related to degree of spoilage - (target value, on the horizontal axis) with that computed by the MLP (vertical axis). The red line and high R value indicate a very good linear fit.

Because a *single* trained MLP is sensitive to the randomly selected initial network weights, 100 training iterations were performed to assess the robustness of this network architecture in this respect. Table 6.1 shows (for all food categories) the mean, median, and standard deviation of the computed R values over all 100 iterations. Clearly, the MLP is better able to track the spoilage of certain categories of foods (e.g., yoghurt) than others (e.g., cream).

Sample Type	Mean R	Std Dev R	Median R
Cream	0.822	0.238	0.893
Eggs	0.949	0.191	0.978
Milk	0.965	0.0367	0.976
Sour Cream	0.933	0.197	0.989
Yoghurt	0.979	0.0584	0.996

Table 6.1: Experiment #1: Mean, standard deviation, and median of correlation coefficients over 100 iterations of MLP training for five food categories.

In this application, it would be ideal if there was a general “spoilage space” to which all of the oldest samples (regardless of category) clustered during spoilage. After all, a wide variety of foods that have “gone off” elicit odours that humans find unpleasant (regardless of origin). In Figure 6.5(a), the complete set of samples (all days, all foods) was used to calculate the PCA projection. With this method, this behaviour is not seen, as the Day 7 samples appear scattered throughout the PCA space. The use of supervised methods, however, may be applicable in this case. An MDA DR method was used wherein the sample *ages* (Day 1, 2, 3, 5, and 7) defined the categories used to train the transformation matrix. By doing so, the system attempts to find features that emphasize the differences between the different spoilage phases (including all foods), in effect learning the feature combinations that distinguish the categories based on age. Upon inspection of Figure 6.5(b), it is evident that the oldest (i.e. most spoiled) samples clearly separate from the rest.

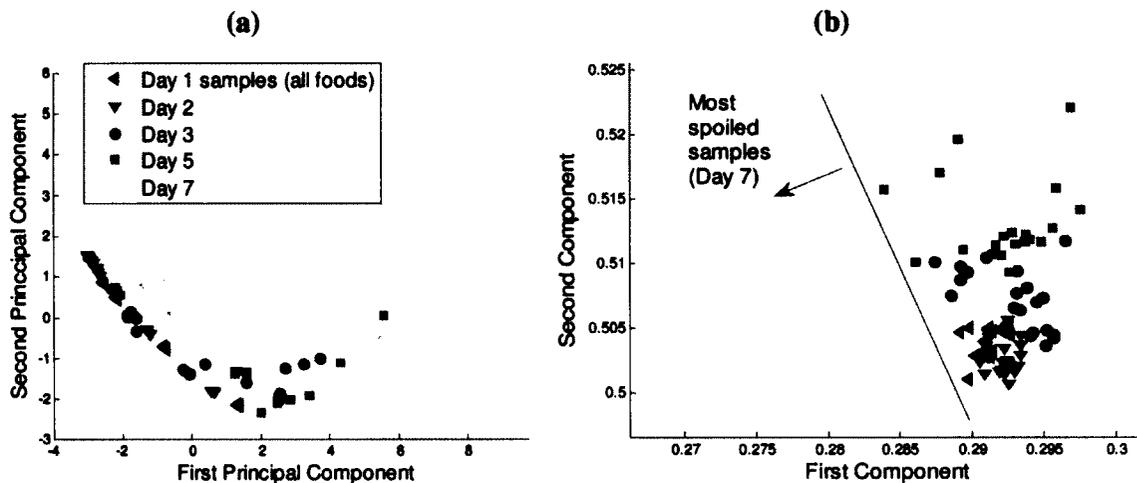


Figure 6.5: Experiment #1: Representation of all food samples – all days, all categories (100 samples): (a) PCA, (b) MDA. In the PCA plot, there is no trend towards a “spoilage” space as the experiment progresses (e.g., the oldest samples overlap with the earlier ones). However, when using MDA (a supervised DR method), the projection into the new space clearly separates the oldest (i.e. most spoiled) food samples.

6.3.4 Experiment #1: Conclusions and Discussion

With reference to Figure 6.3, it is evident that as the experiment progressed, more within-category variability was seen (the clustering of Day 7 samples is much larger than those taken earlier). This is likely related to the sampling process. At the beginning of the experiment, all samples had a low viscosity. This made pipetting very easy and ensured that the volume transferred to the vial was very close to 1.0mL. Towards the end of the week, however, several foods (particularly sour cream and yogurt) had thickened significantly which made pipetting more difficult. It was impossible to confirm that exactly 1.0mL was transferred, as some sample remained adhered to the inside of the pipette tip. Furthermore, as time progressed, the foods were seen to grow moulds, develop crusts on the surface, and separate (See Figure 6.1). In these cases, any

remaining inhomogeneity (after mixing) in the sample would result as variations in the e-nose responses.

The work associated with Experiment #1 has been disseminated in [25].

6.4 Experiment #2 – Validation Study with Single Food and Regular Sampling

6.4.1 Experiment #2: Food Samples, Sample Preparation and Sampling Protocol

Unlike the previous experiment, the intent of Experiment #2 was to provide a more detailed picture of the time evolution of odour signatures by performing more frequent sampling of the food over time. E-nose responses were recorded for a single substance (milk with 3.25% MF) by sampling *regularly* (at 4h intervals) during a week-long period during which the milk was left to spoil at room temperature. Consecutive responses were from *separate* 1mL milk aliquots in 10mL capped vials – a total of 43 responses were recorded over a period of $43 \times 4\text{h} = 172\text{h}$ (just over 7 days).

6.4.2 Experiment #2: Additional Details

Headspace generation parameters were the same as those used in Experiment #1 (300s at a temperature of 50°C), and feature vector construction was performed similarly (fractional baseline manipulation followed by extraction of the maximum absolute value from each sensor curve, and vector normalization).

6.4.3 Experiment #2: Results

Figure 6.6 shows the sensor responses in the PCA space. The freshest samples (#1 – #15) are clustered tightly, indicating very little odour changes in the first 2.5 days or so. After this, the samples follow a general path (to the right with increasing age), but the responses also become increasingly variable.

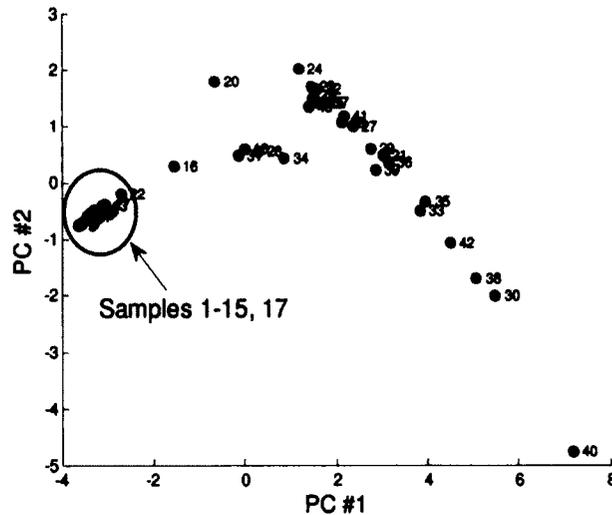


Figure 6.6: Experiment #2: Plot of first and second principal components, for milk sampled for one week at 4h intervals. The sample number next to each point represents the chronological sequence (i.e. sample #1 taken at $t = 0\text{h}$, sample #2 taken at $t = 4\text{h}$, sample #3 taken at $t = 8\text{h}$, etc.).

Unsurprisingly, the increased variability for older samples seen in this experiment made it more challenging for the MLP (with the same architecture and training parameters as in Experiment #1) to track the spoilage patterns over time, as shown in Figure 6.7(a). The mean/std dev/median R value was, respectively, 0.882/0.128/0.911 when computed over 100 training iterations. The FCM clustering results are shown in Figure 6.7(b). The memberships μ_{ic} for each sample i in each cluster c are indicated on the left (spoiled) and right (fresh) axes labels. In this plot, $C = 2$ was specified, which corresponded to the simplest approach (one cluster for fresher foods, the other for spoiled foods). It can be seen from this plot that the fresher milk (e.g., from the first 2-3 days,

samples 1-15) have a high degree of membership ($u > 0.95$) in the “fresh” cluster. Likewise, the oldest milk (from days 5-7, samples 24-43) have been associated with the “spoiled” cluster, though not as prominently ($u > 0.65$). The remaining samples (those processed mid-week) do not consistently associate with either of the clusters.

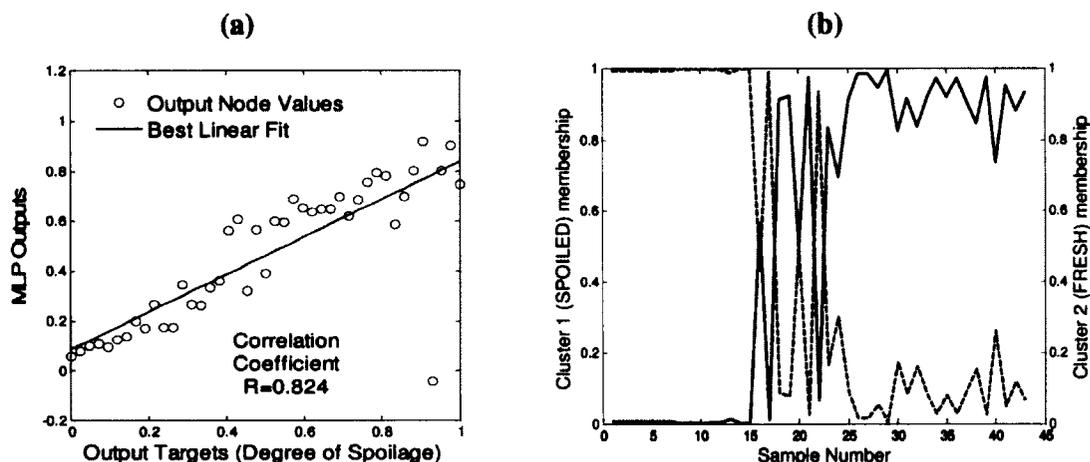


Figure 6.7: Experiment #2: (a) Results of regression analysis comparing the sample age - related to degree of spoilage - (target value, on the horizontal axis) with that computed by the MLP (vertical axis). Notice the difference in sampling in this plot (regular 4h intervals) vs. Figure 6.4. (b) FCM results ($C = 2$). On the horizontal axis, sample number represents the chronological sequence (i.e. sample #1 taken at $t = 0$ h, sample #2 taken at $t = 4$ h, sample #3 taken at $t = 8$ h, etc.). The vertical axis on the right indicates degree of membership in the FRESH cluster, on the left indicates degree of membership in the SPOILED cluster.

6.4.4 Experiment #2: Conclusions and Discussion

At first glance, the variability in the e-nose responses shown above seems unanticipated – a smooth trajectory in the PCA space as the sample spoils would be expected. Recall, however, that in this experiment consecutive sniffs were performed on *different 1mL aliquots of milk in different vials*. It is this fact (as opposed to inherent variability in the spoilage process) that is likely causing the observed high variability during the latter periods in the experiment (days 3-7). One possible explanation is that each (non-sterilized) vial would have contained a different number and type of micro-organisms before the milk was added and the vial capped. Towards the end of the week,

the amount of microbe growth at room temperature in different vials likely led to the odour variations seen. This effect would not have been as prominent in the first couple of days (hence, the tight clustering of samples on the first couple of days before this could occur).

The work associated with Experiment #2 has been disseminated in [27].

6.5 Experiment #3 – Validation Study with Regular Sampling from Same Vial and Varying Storage Temperatures

6.5.1 Experiment #3: Food Samples, Sample Preparation and Sampling Protocol

In contrast to Experiment #2, this experiment ensured that consecutive sniffs taken during spoilage used the *same vial containing the same aliquot* (in an effort to reduce the variability shown above). This experiment also incorporated several additional investigational components:

- The use of *multiple vials* containing the food substance. This was done in order to investigate the repeatability of the odour measurement system. This experiment processed three separate vials of both milk 3.25% MF (labeled M1, M2, M3) and yoghurt with 6% fat (labeled Y1, Y2, Y3).
- The effect of *varying storage temperature*. It is well known that lowering food storage temperatures preserves freshness and arrests spoilage mechanisms [32]. Demonstrating this effect with odour patterns recorded from the same food under different storage conditions provides a degree of validation that the changes seen are due to underlying spoilage processes (in the absence of a confirming microbiological analysis).

As in Experiment #2, regular sampling was performed; each individual vial containing 1mL of food in a 10mL capped vial was sampled every 12h over an 8-day period. The 12h sampling interval was required in order capture responses from all six vials (M1-3 and Y1-3) in each measurement cycle. All six vials were loaded into a temperature-controlled sample tray where they were maintained at a constant temperature (see below) during the measurement period prior to processing. The e-nose sampled the headspace from the vials in an alternating manner (M1, M2, M3, Y1, Y2, Y3). This process was then repeated for each storage temperature considered – 20°C, 25°C and 30°C. These temperatures were selected to ensure spoilage within the week-long measurement period. A total of 276 sniffs were measured, as shown in Table 6.2.

Milk at 20°C	3	16	48
Milk at 25°C	3	16	48
Milk at 30°C	3	14	42*
Yogurt at 20°C	3	16	48
Yogurt at 25°C	3	16	48
Yogurt at 30°C	3	14	42*
			Total = 276

Table 6.2: Experiment #3: Summary of the number of samples used and smellprints generated during this experiment. *: The e-nose stopped unexpectedly towards the end of the 30°C sequence, accounting for the slightly smaller number of sniffs at this temperature.

6.5.2 Experiment #3: Additional Details

The headspace generation parameters used in this protocol were less aggressive than in the previous two experiments – 35°C for 60s (as opposed to 50°C for 300s). These settings were changed for a couple of reasons: a) to better mimic ambient monitoring (as close as possible using the AlphaMOS HS100 – 35°C / 60s were the minimum agitation temperature/time permitted by the software), b) to minimize any destructive effect that

reheating the sample would have (since this protocol required resampling from the same vial). Like the previous experiments, the maximum absolute response from each sensor curve was retained for the e-nose response feature vector (but in contrast to the previous experiments, neither baseline manipulation nor vector normalization was performed).

6.5.3 Experiment #3: Classification

Ultimately, if an e-nose system is to be deployed in a smart home for identification of food unsafe to ingest, it should be capable of learning the relevant features. Accordingly, this experiment also assessed the ability of a classifier to discriminate fresh from spoiling food samples (based on age). To do this, the sniffs generated by the e-nose were categorized into one of three classes according to the time elapsed since the beginning of the experiment:

- **Class A:** Food samples processed on Days 1, 2 and 3. These (samples #1-6) represent the freshest.
- **Class B:** Food samples processed on Days 4 and 5 (samples #7-10).
- **Class C:** Food samples processed on Days 6, 7 and 8. These (samples #11-16) represent those that had been stored the longest (thus more prone to spoilage). Note that in the case of the samples stored at 30°C, there were no vials processed on Day 8 (see Table 6.2).

This experiment employed an LDA classifier (see Section 3.2.5) and to reduce the potential for overtraining, the following splits between training and test data were used:

- **Leave-one-out (LOO) cross-validation** – See Section 3.2.5.

- **Stratified k -fold cross-validation** – With this approach, the data is partitioned into k groups. One is used to test the classifier, and the remaining $(k - 1)$ groups are used to train the classifier. This process is repeated k times (yielding an accuracy of CA_i for that iteration) and the overall CA is averaged over all k iterations. Since three independent samples for each substance were collected, the different vials were used to define the $k = 3$ partitions (e.g., Iteration #1: Training Data consist of the sniffs from Vials M1, M2; Test Data are from Vial M3, etc.). This arrangement has the additional benefit of ensuring the same proportion of samples from each category is represented in the training and test sets (stratification).

There was no DR performed prior to classification (i.e. all 12 input features to the classifier).

6.5.4 Experiment #3: Results

Figures 6.8 and 6.9 reveal that the revised sampling method successfully reduced the variability evident in Experiment #2 for yoghurt and milk, respectively. The time evolution of the feature vector components is shown throughout the 8-day measurement period for yogurt and milk, respectively (at two storage temperatures). For each of the twelve sensors (with the exception of the LY2/LG sensor, which was deemed to be faulty), a relatively smooth transition in the feature vector components is evident as the food ages and spoils – the e-nose sensors appear to track the changing odour characteristics during spoilage without unexplained jumps or discontinuities. Since

similar results are seen when aggregated, there is a high degree of repeatability in this protocol.

When comparing the effect of different temperatures on odour responses, it is evident that the higher temperature (red lines in Figures 6.8 and 6.9) caused a rapid change in the feature vector components (there is a pronounced change around $t = 50\text{h}$, roughly 2 days after the start). The samples stored at the cooler temperature (in blue) do not behave similarly, instead remaining roughly constant for the entire measurement period (a few sensors are just starting to trend significantly by this time). This behaviour makes intuitive sense because spoilage mechanisms (e.g., bacterial growth) are known to accelerate at higher food storage temperatures [32] (though in this study this has not been confirmed with a bacteriological analysis). A similar effect has been reported in other work (e.g., [73]).

Figure 6.10 shows the PCA results for the samples stored at 25°C – in this plot, the mean feature over the three vials is used to calculate the PCA transformation matrix, (i.e. PCA was calculated with 16 (and not 48) input feature vectors). The spoilage trajectories (indicated with arrows) are seen to be unique for each substance.

The MLP results and FCM results also demonstrate an improvement using this sampling protocol. Using the MLP regression approach, there was an improvement in the calculated R value (over 100 training iterations) to 0.936/0.127/0.969 (mean/std dev/median) – results reported for an arbitrarily selected milk sample (M1) – and 0.950/0.126/0.969 – for an arbitrarily selected yoghurt sample (Y1). The degrees of membership in the FCM results (Figure 6.11) also illustrate a much smoother transition between fresh and spoiled categories (as compared to Figure 6.7(b) – Experiment #2).

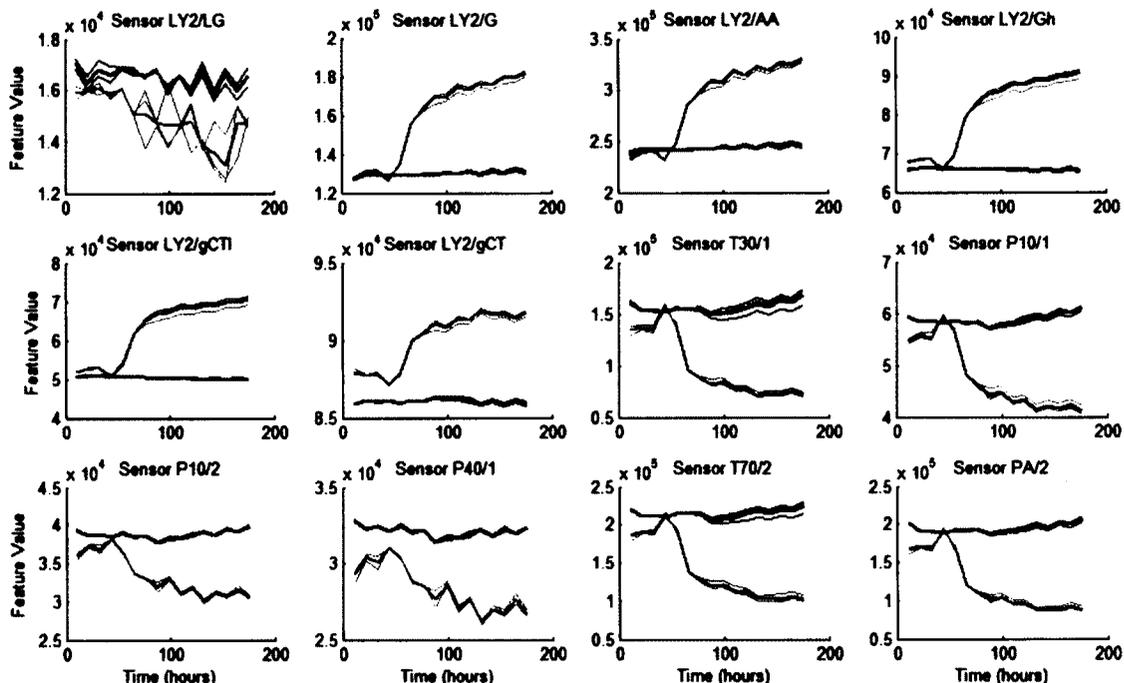


Figure 6.8: Experiment #3: Time progression of feature vector components for yogurt. There is one subplot for each MOS sensor and the sensor name is labeled in the title of each. The thin lines represent feature values for each individual vial (Y1, Y2, Y3) – the mean over all vials is shown in a larger line width. Both storage temperatures are represented in each subplot (20°C: blue, 25°C: red). Sensor LY2/LG behaved erratically throughout the experiment.

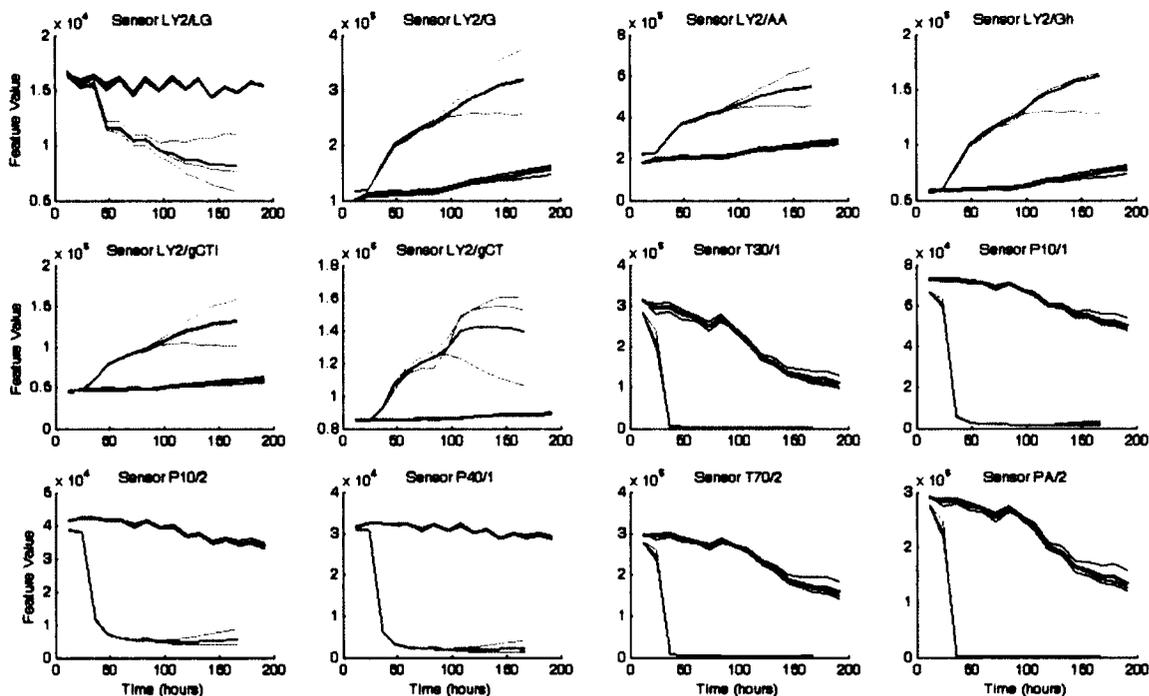


Figure 6.9: Experiment #3: Time progression of feature vector components for milk. There is one subplot for each MOS sensor and the sensor name is labeled in the title of each. The thin lines represent feature values for each individual vial (M1, M2, M3) – the mean over all vials is shown in a larger line width. Both storage temperatures are represented in each subplot (25°C: blue, 30°C: red).

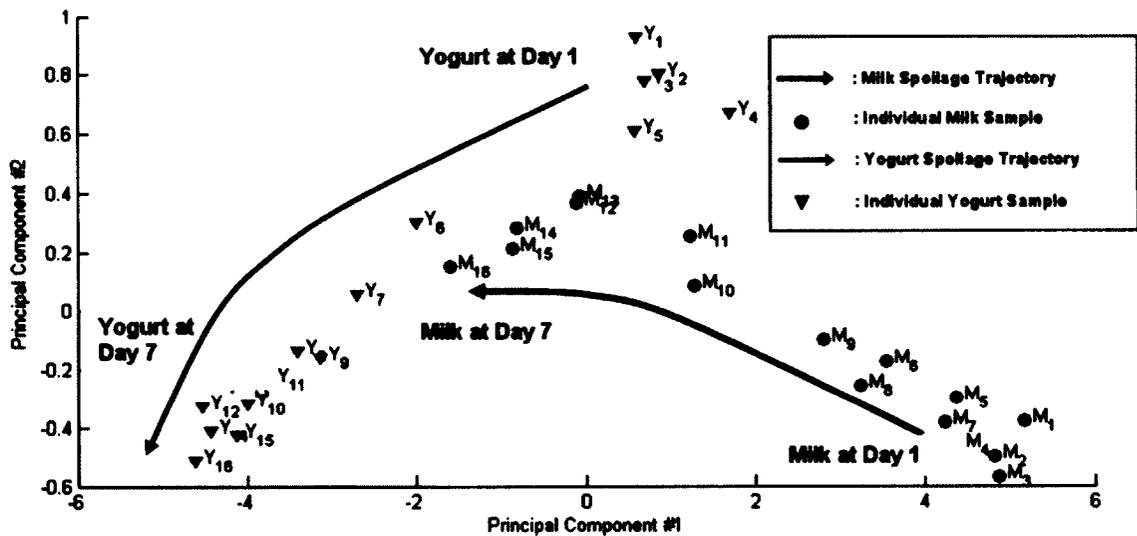


Figure 6.10: Experiment #3: PCA representation of milk and yogurt e-nose responses (computed with an average over all three vials). The subscript indicates the chronological number in the sequence, e.g., M_1 is the first milk samples ($t=0$), M_2 is the second milk samples ($t = 12h$), M_3 is the third milk samples ($t=24h$), etc. Y (red triangles) = yogurt, M (blue circles) = milk. The arrows indicate the general trajectory of the samples in the PCA space as the foods spoil.

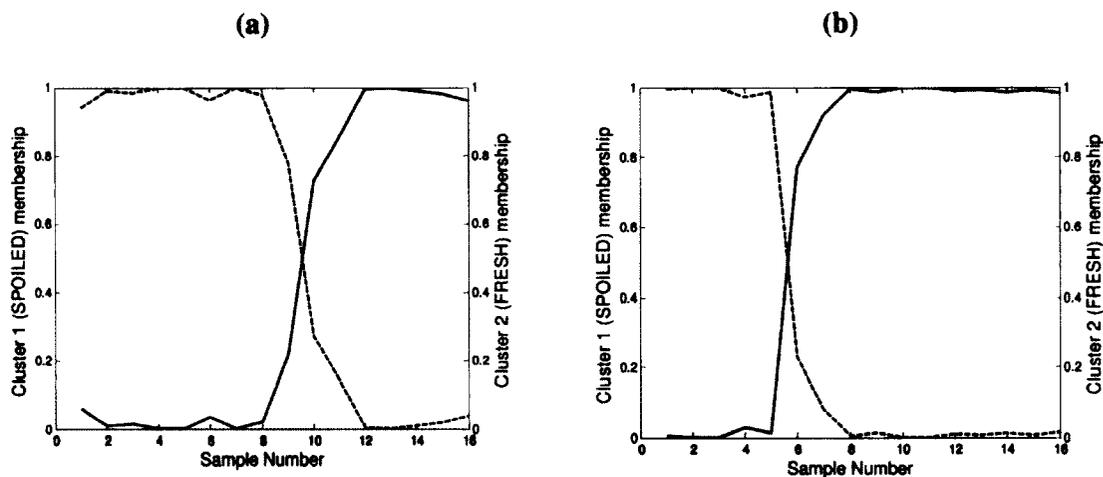


Figure 6.11: Experiment #3: FCM results ($C = 2$) for a) Milk, b) Yoghurt. On the horizontal axis, sample number represents the chronological sequence (i.e. sample #1 taken at $t = 0h$, sample #2 taken at $t = 12h$, sample #3 taken at $t = 24h$, etc.). The vertical axis on the right indicates degree of membership in the FRESH cluster, on the left indicates degree of membership in the SPOILED cluster.

The results of classifying the food samples based on age is shown in Table 6.3 for each food and storage temperature, using the validation methods described above.

Classification accuracies are consistently higher than 85% using LOO (and in some cases > 95%), demonstrating the system's ability to distinguish food samples based on age for all foods and storage temperatures tested in this experiment.

Milk at 20°C	89.9 %	68.8 / 93.8 / 87.5	83.3 %
Milk at 25°C	91.7 %	87.5 / 100 / 75.0	87.5 %
Milk at 30°C	90.5 %	100 / 57.1 / 64.3	73.8 %
Yogurt at 20°C	97.9 %	100 / 100 / 100	100 %
Yogurt at 25°C	97.9 %	100 / 100 / 87.5	95.8 %
Yogurt at 30°C	100 %	100 / 100 / 57.1	85.7 %

Table 6.3: Experiment #3: Classification accuracies for food and storage temperature combinations. For *k*-fold validation results, the per-fold CA is shown using (left-to-right): (Train Vials 1&2, Test Vial 3) / (Train Vials 1&3, Test Vial 2) / (Train Vials 2&3, Test Vial 1). Classification was performed using 12 input features (no DR).

6.5.5 Experiment #3: Conclusions and Discussion

In this experiment, the decision to focus on relatively high storage temperatures (20°C, 25°C, and 30°C) was motivated by the necessity to complete the sampling within a reasonable time frame (roughly one week). In exploratory work (results not shown), the identical sampling protocol was performed at 4°C – a typical refrigeration temperature – but the e-nose features did not change appreciably within a 9-day period. By using the higher storage temperature instead, it allowed us to demonstrate the effect of varying temperatures on the odour patterns as the foods aged. Verification that the e-nose odour patterns demonstrate a similar trend over a wider range of storage temperatures (with longer storage times) should form the basis of future work.

Figures 6.8 and 6.9 demonstrate that overall, the inter-vial repeatability of the sensor responses was quite good, the exception being milk at a storage temperature of 30°C (see

red curves in Figure 6.9). This combination showed a larger variation in the trajectory of the features as the samples aged. This is likely due to the manner in which sampling was performed. Sterile vials were not used, so even though vials were capped at $t = 0$, the spoilage rate may vary based on any organisms/bacteria initially present in the vial (as was hypothesized earlier in Experiment #2). These results would suggest that this effect is exacerbated at higher temperatures.

Classification was performed by dividing the sniffs into three categories, based on time. The selected cut-off points for these categories were arbitrary, though care was taken to represent each of the three categories with roughly the same number of samples. Further testing (results not shown) confirmed that the classification results were not that sensitive to the time thresholds selected to distinguish categories. In the absence of a bacteriological analysis, it is impossible to ascertain the actual degree of spoilage (e.g., due to bacterial growth) – duration of storage time is not a perfect correlate. Given this, future work will further investigate the use of fuzzy methods, which allows *graded membership* in different categories (e.g., fresh/questionable/spoiled). Use of these techniques in a similar domain (food quality assessment with an *electronic tongue* sensor array) has been shown to be effective [134].

A portion of the work associated with Experiment #3 has already been disseminated in [26], and the remaining portion has been submitted in a journal manuscript already mentioned.

6.6 Integration of Odour Monitoring into a Smart Home: Proposed Framework

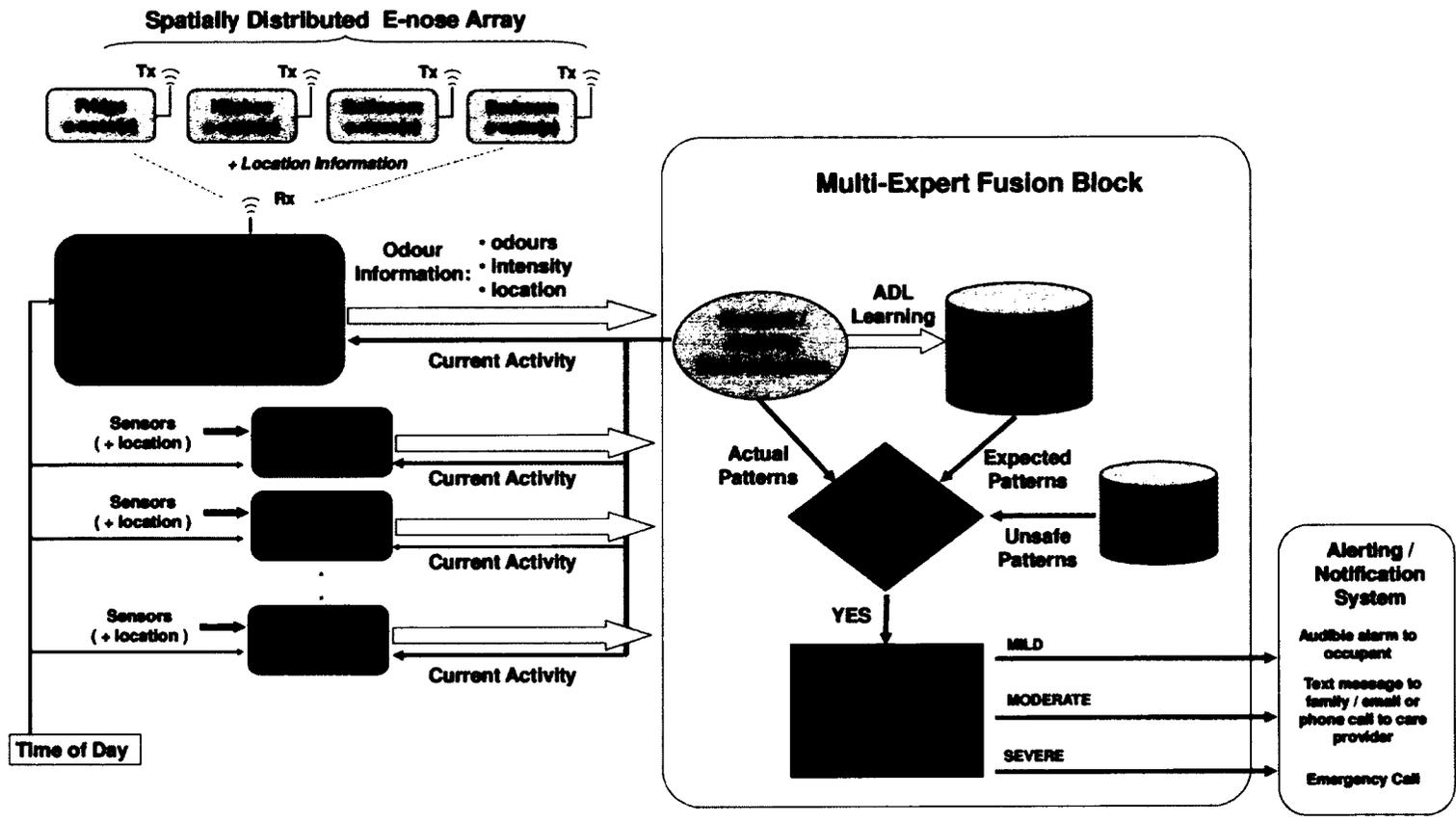
Detection of food spoilage (the problem investigated in the experimental portion of this chapter) is just one potential application of using machine olfaction in the residential environment. Though commercial e-nose systems have been available for over a decade, this sensor modality has not been widely explored in the smart home context, as introduced in Section 6.1. In [135], the authors acknowledge the value of deploying e-nose in the residence. They propose a system that includes: a) a network of simple, location-aware gas sensors that can highlight a potential problem and b) a robot equipped with a more sensitive e-nose to make the final determination.

The human sense of smell is one key way of receiving information about the environment. In a residential setting, people are constantly recognizing smells and associating these with activities or events such as baking, cleaning, dampness, or unhygienic conditions. In a certain sense, then, humans already perform a certain degree of “odour monitoring” (even though this may be a largely automatic or subconscious activity). Given this, it is intuitive to attempt to complement the human sense of smell with a smart home odour monitoring system based on e-noses, which may assist in recognizing odour events that: a) are too subtle for the human olfactory system to detect and/or b) may not be easily detected by a resident due to dementia or olfactory impairment [136].

This section introduces a framework that would be suitable for incorporating e-noses into a smart home monitoring system. The key elements of this framework are illustrated

in Figure 6.12. Briefly, an *array of e-noses* throughout the residence is used to measure local sensor responses, which are sent *wirelessly* to an *e-nose expert system* capable of translating this data into useful information (type of odour, location, intensity, etc.). This is then combined with other smart home sensors (and higher-level information derived from other expert systems) in a *fusion block*. This component aggregates this information, determines the context of activities within the smart home, and maintains a profile of the occupant's activities of daily living (ADLs). When the environment differs significantly from expected (for example, due to a smell of natural gas), the fusion block issues a message to an *alerting system* that would initiate a notification commensurate with the perceived conditions. Note that in this discussion, "single-application" sensors such as carbon monoxide and smoke detectors are deliberately excluded. These are engineered to test for their targets with high sensitivity and specificity, but unlike the e-nose they do not have the capability to be trained to learn other odours. Full descriptions of the key components in this framework appear below.

Figure 6.12: Key elements of a smart home environment system that includes odour monitoring.



6.6.1 Spatially Distributed E-nose Array

A single e-nose would not have sufficient sensitivity and discrimination capability to identify a wide range of odours from all relevant locations in a smart home. For this reason, e-noses are better deployed in an array configuration throughout the residence. At a minimum, odour sensors should be positioned at the key locations: a) in the kitchen, b) inside the refrigerator, c) in the bathroom, d) in the bedroom. This arrangement would also have an additional benefit as follows. If a single e-nose was employed to detect all odour types, classification performance would likely decrease, since the large number of possible odours would likely overwhelm the pattern recognition system. In a distributed array, though, each e-nose would have a unique location (thus, a smaller set of expected odour possibilities to deal with) and correspondingly better discrimination power [135]. This architecture might also enable the possibility of *smell localization*, based on the spatial pattern of e-nose responses [137].

6.6.2 Wireless communication capability

Implicit in the discussion above is the ability for the e-noses to communicate with an expert system. This property is crucial, since it enables the overall integration of odour information (as well as information from other ambient sensors, if deployed). A wireless system based on standard short-range communication protocols such as Bluetooth or IEEE 802.15.4 (Zigbee) [138] would be fitting for this purpose. The Zigbee configuration is easy to deploy, unobtrusive, and its integration in e-nose systems has already been demonstrated [139] [140].

6.6.3 Expert Systems

The utility of the odour monitoring system is enhanced when deployed in a smart home residence already equipped with other smart home sensing modalities (such as those mentioned in Section 6.1). This arrangement can offer additional contextual information that may better guide the necessary response. Each of the sensor modalities has an associated *expert system* that is able to extract higher-level information from the sensor responses. For the e-nose expert system, the effect of several factors that degrade e-nose performance can be mitigated with appropriate digital signal processing (DSP). For example, sensor drift is an inherent property of e-noses which, if not addressed, would require an unacceptable amount of retraining in order to maintain the integrity of the classification engine [84]. Software-based drift compensation methods use DSP methods such as wavelet analysis [141], dimensionality reduction [142] and system identification theory [143]. The e-nose expert system could also take a set of measurements from each e-nose and, after comparing these to its training base and taking location and time components into account, determine that there has a smell of urine present in a particular location for several hours, for example. Similarly, an expert system constructed around bed sensors might be able to identify that the occupant is currently sleeping. In this framework, expert systems should be context-aware in terms of the ability to record *location* and *temporal* information (such as the time, duration and frequency), in order to provide important input to the multi-expert fusion block.

6.6.4 Multi-Expert Fusion Block

All of the information from the individual expert systems is provided to a multi-expert fusion block. This component acts as the system's central intelligence hub, and interfaces

with all components described above. It incorporates artificial intelligence (AI) algorithms and determines the nature of the overall perceived odour environment (in the context of other patterns of events from all expert systems, over both location and time). A database consisting of potentially unsafe events is maintained – for instance, an urgent situation likely exists if a detected burning odour is combined with the information that the occupant has recently fallen (as could be determined with a fall detection expert system, e.g., [144]). Similarly, if a cigarette odour is detected in the bedroom at the same time that bed sensor expert system indicate the occupant is lying down or sleeping [145], a potentially more serious situation exists than if either of these events were identified in isolation. Another important function of this block is the ability to learn the odour events that would be considered normal in the residence (ADLs) –e.g., baking, frequency of cleaning, etc. By doing so, the system would be able to adapt to the individual behavioral habits of the occupant when deployed.

Based on this information, this block must be able to decide whether the odour environment is *normal* or *abnormal* – a determination that is based on input from the expert systems, as well as learned ADLs and potentially unsafe events. As another example, the system should be able to differentiate between an expected, innocuous event (e.g., temporary urine odour in the bathroom) from another of more concern (e.g., prolonged urine odour in the bedroom). If the event is deemed to be abnormal, the fusion block would decide if it is necessary to trigger a system-initiated alert appropriate for the situation.

6.6.5 Alerting System

Any alert must be suitable for the severity of the situation at hand. For example, the occupant can be informed of minor events with a simple audible reminder. Other more serious detected events may merit a notification to the resident's family member or caregiver, using the technology appropriate for that situation (e.g., email notification for moderate severity, or immediate text message/call to mobile phone for high severity) [125]. An important consideration here is the ability to tune the system to the appropriate sensitivity and specificity. Too frequent alarms (i.e. false positives) will compromise the system's utility and risk being ignored or disabled by the occupant. Alarms must also have the ability to be temporarily disabled (e.g., if the underlying reason for a mould smell has been addressed, but the odour is still present in the residence and gradually diminishing).

6.6.6 Practical Issues to Overcome

The term introduced previously – *ambient-assisted* living environment – highlights a key acceptance criterion. The device(s) described above must not be overly inconvenient to use or intrusive to the occupant. To the extent possible, these technologies should “blend in” to the living environment in order to protect privacy and maximize the occupant's comfort with the technology. Presently available commercial e-nose systems (such as the AlphaMOS instrument used in this thesis) are clearly not well-suited for the smart home environment. Some important factors prohibiting their adoption include high cost, large system size, carrier gas delivery mechanism, and the need for manual intervention and sample preparation. Fortunately, there is active research into many of the shortcomings of current industrial e-nose systems that make it presently unsuitable for smart homes. There are several portable e-nose instruments that have been developed, such as the

commercially available Cyranose 320 (described in Chapter 3) and zNose, as well as the Kamina gradient microarray [146] and NASA's 3rd generation Jet Propulsion Laboratory e-nose [147]. The gas sensors are continually being miniaturized, making them more attractive for use in pervasive health monitoring environments. These new smaller sensors are now being fabricated using commodity technologies which will drastically lower manufacturing costs [148][85]. Additionally, e-noses are now being fitted with wireless data transmission features [139], [149], [150]. Once this technology matures, it will be easier to aggregate the odour signals from multiple e-nose devices (positioned in different rooms) to a central odour monitoring station in the smart home. Finally, e-nose systems are being developed that have a much better capability of sampling ambient environments [151][152].

6.7 Overall Summary and Conclusions

This chapter described experiments in which a commercially available e-nose (based on MOS sensors) was used to track odour patterns associated with food spoilage. Three experiments were performed, differing mainly in terms of the method in which the foods were sampled and processed by the e-nose. It was demonstrated that as the foods aged, the smellprints associated with this process followed generally smooth trajectories in the feature space (that were unique for different foods). Because this was motivated by other work in smart home monitoring, a more general framework for a smart home odour monitoring system was then introduced and described. The experimental results described herein represent exploratory work in this area – though promising results were shown, the last section highlights several technical issues that must be surmounted before

this technology becomes practical in this environment. Despite this, there are promising signs that suggest deployment of e-noses in the smart home context will be viable in the future. Detection of food spoilage (the problem investigated in this chapter) is just one potential application in this environment.

With reference to Figure 6.13, this chapter (in contrast to the others) used real food samples (as opposed to bacteria), which are available at the earliest point in the process. The ability of a commercially available e-nose to track the changing odour characteristics of several spoiling foods was demonstrated. Admittedly, this was done without regard to the underlying mechanism (the degree to which bacteria growth or other spoilage process was responsible was unknown). Regardless, this work suggests that the development of a simple e-nose based device capable of providing simple spoiled/not spoiled decisions is conceivable – as part of an *in situ* odour monitoring system (e.g., in a smart home). The experimental results in this chapter suggest that for at least one smart home application – food spoilage – the underlying e-nose sensors (which by themselves are physically small) have sufficient sensitivity or specificity to justify continued research in this area.

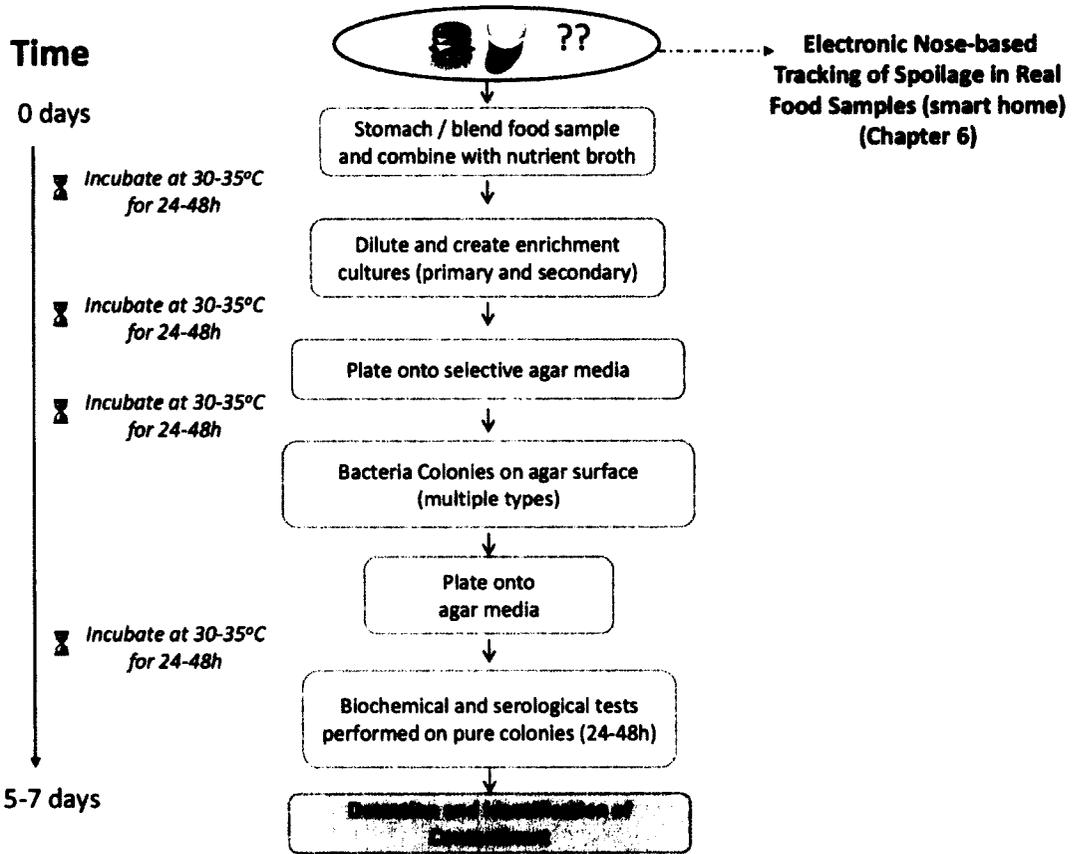


Figure 6.13: Simplified view of conventional methods. Highlighted block indicates the stage at which the sample type considered in Chapter 6 is available.

Chapter Seven

Conclusions and Future Research

7.1 Summary of Research

Chapter 1 provided the motivation for the overall problem addressed in the thesis (an unacceptably high incidence of food borne illness worldwide), followed by an overview of conventional methods for food safety testing and a description of the important shortcomings associated with these methods (long time required for results, high cost and complexity, and lack of portability).

In Chapter 2, a more detailed description of the conventional methods was presented, followed by a review of a number of enhancements to these that seek to improve on the limitations inherent to the traditional approaches. A review of existing biosensor-based systems was provided, with particular attention placed on assessing their suitability for incorporation into the areas of food safety and pathogen detection. Particular attention was given to existing research that used biosensor-based systems with some success in this application area.

The biosensor types examined in this thesis were the *electronic nose* and *Raman spectroscopy*. In order to determine whether these biosensor-based systems held the potential to identify possible food contaminants, experimental work was performed using sample types at various stages in the overall food inspection process (see below). Off-the-shelf biosensors were combined with sensibly chosen signal processing and pattern classification algorithms to form measurement systems that were able to identify contamination and/or discriminate between potential contaminants (e.g., bacteria). In each case, the results reported warrant further research in these areas.

Specific research results (organized by the thesis chapter in which the relevant experimental work was performed) appear below.

7.2 Summary of Research Results: Chapter 3

This chapter described an e-nose based method for discrimination of samples containing bacteria (*E. coli* and *Listeria*) suspended in a nutrient broth of varying concentrations.

- A preliminary analysis with a small number of samples demonstrated that three e-nose instruments of different sensor modality (MOS, FMS and CPS) were able to distinguish the bacteria in a dense culture ($\sim 10^9$ cells/mL), with MOS and FMS outperforming the CPS. (It was suggested that poorer performance of the CPS enose, a handheld instrument, is likely due to the lack of controlled headspace generation (and not necessarily due to the lack of selectivity/sensitivity of the CPS sensors *per se*).
- For a given e-nose experimental run (lasting several hours), MOS and FMS based e-noses were able to retain discrimination between categories when the concentration was lowered from 10^8 cells/mL to 10^6 cells/mL (by serial dilution with additional nutrient broth), but this required the use of a supervised DR technique (ULDA). For the MOS e-nose, the classification accuracy decreased as the bacterial concentration decreased.
- The ability to discriminate between bacteria categories is lost when using samples captured over multiple days within a one week period. This behaviour was shown to be due to significant drift in the odour signature of the dilution medium.

- Adjusting the sampling protocol to mitigate the nutrient broth drift effect (by preparing all dilutions at the outset) suggested discrimination between bacteria is feasible at concentrations lower than 10^6 cells/mL.

7.3 Summary of Research Results: Chapter 4

This chapter described a MOS e-nose based method for discrimination of samples containing single colonies of bacteria (using PBS as a suspension medium), thereby eliminating the drift-prone nutrient broth component from the sample mixture.

- An initial feasibility experiment using 105 samples (*E. coli*, *Listeria* and control group) sampled over a week-long period yielded classification accuracies of 92.4% (when used in conjunction with supervised DR, a linear classifier, and leave-one-out cross validation).
- A further study with four bacteria types validated these results ($CA = 92\%$), while simultaneously controlling for the effects of differing agar substrate and colony size. This work represents the first time that discrimination between the odour patterns of individual bacteria colonies has been reported using an e-nose system.
- Use of multiple consecutive sniffs on each colony increased classification accuracy (by eliminating the conditioning effect of the first sniff and combining all others into a composite response).
- Use of PBS as a suspension medium did not pose the same time-varying drift effect as was evident with nutrient broth (Chapter 3).
- Because this sample type is available relatively early in the food inspection process, it might find utility in a pre-screening role. It was demonstrated that the

degree of trustworthiness in the classifier output for each sample can be measured with confidence metrics based on either posterior probabilities or silhouette coefficients. By applying a threshold, increased accuracy is reported for those samples that are deemed to lie outside the classifier's "reject region" (which defines those samples deemed to be unclassifiable).

7.4 Summary of Research Results: Chapter 5

This chapter described a SERS-based method for discrimination of *Listeria* bacterial species (in a silver colloidal mixture), a more challenging problem than genera-based bacteria discrimination described in Chapters 3 and 4. Raman spectra were measured from a portable spectral acquisition unit suitable for *in situ* deployment (developed elsewhere).

- An experiment using 483 spectra from six *Listeria* species (including the human pathogen *L. monocytogenes*) demonstrated a classification accuracy of 96.1%. This work represents the first time that discrimination between *Listeria* species using a SERS-based system has been reported.
- The descriptive features that enabled this discrimination were based on the stationary wavelet transform (SWT), a multiscale representation of the input SERS spectrum. The SWT was used for noise removal, baseline removal and feature vector extraction based on the location of significant SWT coefficient values, followed by dimensionality reduction using MDA.
- In addition to LOO cross-validation, bootstrap resampling was used to estimate classification accuracy.

- Confidence in classifier outputs were calculated using the same methods in Chapter 4, in an attempt to screen out those samples that are not identified with sufficient dependability using this technique (which could also be used in a pre-screening capacity).

7.5 Summary of Research Results: Chapter 6

Deviating from previous chapters (wherein the samples were based on pure bacteria), Chapter 6 reported results measuring the evolution of odour patterns associated with the spoilage of real food samples. Food spoilage is one example of an adverse scenario that might develop in a smart home. More generally, the use of e-nose technology could introduce the ability to identify other odour events (including ADLs, but also those indicative of a problem) in a smart home. This work represents the first time that the viability of using odour sensing technologies has been assessed in a smart home setting (with particular attention as an assistive technology for older adults).

- There were smooth trajectories evident in the odour patterns measured from a wide variety of foods left out to spoil over a week-long period. This was evident in the raw sensor feature space as well as in the DR feature space. Data analysis tools employed in this analysis included PCA, MLP and FCM.
- Measured odour patterns were dependent on the manner in which the samples were prepared and stored prior to e-nose sampling. The best repeatability in results was demonstrated when odour responses were captured from a single vial, and foods stored at the lower temperatures exhibited less change in the odour patterns during the spoilage period.

- These experiments used a commercial MOS e-nose (admittedly unsuitable for smart home deployment). Chapter 6 contains a description of the current shortcomings with e-nose instruments that would currently restrict their adoption in this environment.
- The latter portion of this chapter proposes a novel framework suitable for incorporating an odour sensing system into a smart home with other assistive sensor modalities. Practical considerations such as the number and position of e-noses, integration and communication with other systems and multi-expert fusion are addressed.

7.6 Conclusions

As stated in Section 1.3, the objective of this thesis was to investigate the feasibility and potential of using biosensor-based measurement systems as attractive alternatives to the conventional methods used in food safety. As a result of the work performed in the thesis (and summarized above), the following conclusions can be made:

- An e-nose based biosensor system (when coupled with appropriate signal processing and classification components) was able to discriminate between certain non-pathogenic bacteria at the genus level, using bacteria samples available relatively early in the inspection process. The system's ability to differentiate between bacteria was compromised at lower concentration.
- Using single colonies of bacteria (combined with saline) as the input to the e-nose analysis outperformed the use of bacteria in nutrient broth (which was prone to drift).

- With the e-nose based methods, repeated sampling of bacteria has the potential to improve classification performance (by discarding the first conditioning sniff and forming a composite from the remaining responses).
- A surface enhanced Raman spectroscopy based biosensor system (with signal processing and feature extraction components based on the wavelet transform) was able to discriminate between species of *Listeria* bacteria using bacteria samples available relatively early in the inspection process.
- An e-nose system was able to track the changes in odour associated with the spoilage of several real food samples, and the e-nose sensor responses varied with the storage temperature of the samples (higher temperature led to earlier changes).
- Beyond food safety, there are interesting applications of odour monitoring in a smart home setting which can be explored using the multi-modal framework proposed.

It is very important to note that the biosensor alternatives explored in the thesis are not intended to displace the conventional methods, which will represent the gold standard in food safety testing for the foreseeable future. For these conventional methods, the mechanisms are well understood and their accuracy is near perfect. Though the initial assessment of the accuracy of the systems explored in this thesis (>80%) were promising, it is most likely that a biosensor-based method will not rival the conventional methods in this respect. However, the results demonstrated in the thesis indicate that it is sensible to further investigate these biosensor systems in instances where the advantages they present (potential for expedited results, lower relative cost, ease of operation and portability) are deemed to be important to the application.

7.7 Suggestions for Future Research

Although the experimental work reported in this thesis has demonstrated that the use of e-nose and Raman spectroscopy holds a significant degree of promise in the area of food safety, there are several suggested topics of further investigation that might guide future researchers in this area. These are summarized below:

- **Validation of Methods with Larger Selection of Bacteria (including Pathogenic Bacteria)**

Most of the experimental work in this thesis (with the exception of the Raman spectroscopy work in Chapter 5) was performed on *non-pathogenic* bacteria. This was out of necessity because the e-nose laboratory at Carleton University is equipped to handle bacteria at Risk Group 1 and not higher [153]. Obviously, however, organizations such as CFIA are more interested in human pathogens (such as *E. coli* O157:H7, *Salmonella typhimurium* and *Campylobacter jejuni*). In order to ascertain that discrimination amongst these is also possible, the methods in Chapter 4 (single colony) should be validated in a facility with the appropriate biosafety level. Further work might also measure e-nose responses from a much wider range of bacteria and thereby develop a large library of odour signatures. Methods developed to discriminate between all of these categories must be validated with a large independent test set (i.e., using samples not encountered during training).

- **Investigate Use of Alternate Feature Extraction Methods**

For the experiments described in this thesis, the features extracted from the MOS response curves (the use of the maximum absolute sensor response for each curve, as described in Section 3.2.3.2) were consistent with the conventional techniques that are predominant in the e-nose literature and provided encouraging results. Though the exploratory work that was performed to investigate the value of other feature extraction methods did not enhance classification results significantly in this case, this direction of investigation should not be abandoned entirely. Each MOS response curve provides additional information that might result in a more information-rich feature vector using alternatives features extracted from either the time domain (as was described with the derivative-based measure in Section 4.2.3) or from alternative representations (e.g. model fitting, DWT).

- **Investigate Use of Other Classifiers**

The results reported in this thesis were based on the use of an LDA classifier, a relatively simple choice that generates linear decision boundaries between classes in the feature space. This decision was deliberate – it was deemed appropriate in this exploratory work to focus on the extraction of meaningful features from the biosensor responses. Classification results were reported, but these were intended to illustrate the overall potential of the method – the choice of LDA was sufficient for this purpose. There are a whole host of more sophisticated classification methods available – among these include support vector machines, artificial neural networks (both MLP and radial basis

functions), self-organizing maps, k -nearest neighbours, and others [86][90]. As research continues in this area, the increased sensitivity and specificity required by practical food inspection methods will require the performance gains that could be expected from these alternatives.

- **Incorporate Drift Compensation Techniques into Methods**

As mentioned throughout the thesis (e.g., Section 6.6.3), e-nose instruments are known to suffer from a long-term sensor drift which, if not addressed properly with drift compensation techniques, would require frequent re-training of the system. This problem was not addressed in the thesis, since all data collection for bacteria samples were largely done “point-in-time” (i.e. over a relatively short duration). The baseline sensor responses were shown to be roughly constant using software in the AlphaMOS FOX that demonstrated return to baseline values throughout the sampling sequence. The importance of the drift effect cannot be overstated, however – any system that might ultimately be deployed in the field based on the methods described herein must address this issue.

- **Exploit Multi-modal Inputs and Alternative Sensors to Increase Performance**

In this thesis, a variety of biosensor modalities were used, including MOS-, FMS-, and CPS-based electronic noses and Raman spectroscopy. The discrimination potential of these instruments was reported individually. It is widely known that the use of multiple sensor modalities can improve the overall performance of a classification system in cases where the different sensors provide complementary information [154][155]. There is

every reason to believe that combining the outputs of multiple instruments into a single feature vector would result in superior performance (both in terms of classification ability and determination of detection thresholds). As mentioned in Chapter 6, fusion might also be performed at the classifier level to further improve performance.

Finally, it should be recognized again that the results reported were achieved with MOS sensors recommended by the vendor for generalized applications. In the future, further consultation with e-nose vendors and sensor development groups may reveal an optimized sensor selection (better suited for discrimination of VOCs from bacteria).

- **Investigate Potential to Incorporate Ambient Odour Monitoring Systems into Smart Home**

The experimental work in Chapter 6 was performed at a time when e-nose instruments capable of sampling ambient environments were not available (the only alternative available, the Cyranose 320, became defective after the work in Chapter 3 was performed). However, at the time of preparation of this thesis, several instruments appropriate for this type of sampling were starting to emerge (both research and commercial). It is recommended that future researchers in this area consider these instruments for possible adoption into the TAFETA Smart Apartment. By doing so, it will be possible to perform research that assesses whether identification of “odour events” (some of which were suggested in Chapter 6) – possibly in conjunction with other sensor modalities – is a viable alternative for identifying real problems in this environment.

- **Extension to Other Applications (e.g. Medical Diagnostics)**

The overall problem of pathogen identification is not limited to applications of food safety. As mentioned previously (Section 2.4.2), medical diagnosis using e-nose technology is a very active research area. For example, there are reports of the ability to use e-nose responses as a way of identifying sepsis (from blood cultures) [95], lung cancer (from exhaled breath) [156] and tuberculosis (from sputum samples) [157]. The methods developed in this thesis (particularly those described in Chapter 4) are applicable to certain samples of a biomedical origin (e.g., cultures originating from throat swab). Future work in the biomedical area (particularly for point-of-care applications) will have to surmount the problem of dealing with very complex mixtures (e.g., exhaled breath).

- **Validation of Food Spoilage Work**

The results in Chapter 6 demonstrated that there were marked changes in the odour patterns of spoiling foods. This work can be extended in several ways (see Section 6.5.5). Measurements over a wider range of storage temperatures should be performed (lower than the 20-30°C range investigated in Chapter 6). This was attempted multiple times, but unfortunately the e-nose crashed frequently during the later stages of these longer spoilage times (> 1 week), rendering it impossible to collect the results. Ideally, this activity would also be done in the context of a larger range of food samples (e.g., meat, fish) to establish the generalization potential of the method.

- **Further Development of Confidence Measures**

In all likelihood, if the adoption of e-nose technology into the food safety realm is to occur, it will appear as a complementary technology (it will not replace the conventional methods outright). It might find its first use in a pre-screening function (as has been mentioned at various times throughout the thesis). Accurately determining the confidence in e-nose derived classification results is imperative, and in the thesis a couple of approaches to do this have been investigated. Since many other methods exist (e.g., determining the distance of an unknown sample to the decision boundaries in the feature space, thresholding the posterior probability estimates that result from *other* classifiers (besides LDA) [97]), further work on how these might be used (possibly in combination with one another) is merited.

- **Further Use of Composite Sniffs**

The advantage of combining consecutive sniffs (and eliminating the first one) on classification performance was shown using single bacteria colonies (Chapter 4). The same idea, however, could be applied to the other types of samples used in this thesis (e.g., food spoilage), as long as the application does not require too frequent a sampling rate. Furthermore, more sophisticated combinations of the retained sniffs (beyond the simple average used in Chapter 4) could likely result in better performance.

References

- [1] "WHO - Food Safety and Foodborne Illness." [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs237/en/>. [Accessed: 04-Dec-2011].
- [2] P. S. Mead et al., "Food-related illness and death in the United States.," *Emerging Infectious Diseases*, vol. 5, no. 5, pp. 607-625, 1999.
- [3] T. Roberts, "WTP estimates of the societal costs of US food-borne illness," *American Journal of Agricultural Economics*, vol. 89, no. 5, pp. 1183-1188, 2007.
- [4] J. C. Buzby and T. Roberts, "The economics of enteric infections: human foodborne disease costs," *Gastroenterology*, vol. 136, no. 6, pp. 1851-1862, 2009.
- [5] "WHO - The World Health Report 2007 - A Safer Future: Global Public Health Security in the 21st Century" [Online]. Available: <http://www.who.int/whr/2007/en/index.html>. [Accessed: 04-Dec-2011].
- [6] J. L. Smith, "Foodborne illness in the elderly," *Journal of Food Protection*, vol. 61, no. 9, pp. 1229-1239, 1998.
- [7] P. A. Kendall, V. Val Hillers, and L. C. Medeiros, "Food Safety Guidance for Older Adults," *Clinical Infectious Diseases*, vol. 42, no. 9, pp. 1298 -1304, May 2006.
- [8] S. Mortimore and C. Wallace, *HACCP: A practical approach*, Chapman and Hall, 1998.
- [9] J. G. Surak, "A recipe for safe food: ISO 22000 and HACCP," *Quality Progress*, vol. 40, no. 10, pp. 21-27, 2007.
- [10] V. Velusamy, K. Arshak, O. Korostynska, K. Oliwa, and C. Adley, "An overview of foodborne pathogen detection: In the perspective of biosensors," *Biotechnology Advances*, vol. 28, no. 2, pp. 232-254, 2010.
- [11] A. K. Deisingh and M. Thompson, "Biosensors for the detection of bacteria," *Canadian Journal of Microbiology*, vol. 50, no. 2, pp. 69-77, 2004.
- [12] E. C. Alcocilja and S. M. Radke, "Market analysis of biosensors for food safety," *Biosensors and Bioelectronics*, vol. 18, no. 5-6, pp. 841-846, 2003.
- [13] R. Teratanavat and N. H. Hooker, "Understanding the characteristics of US meat and poultry recalls: 1994-2002," *Food Control*, vol. 15, no. 5, pp. 359-367, 2004.
- [14] Food and Agriculture Organization of the United Nations, "Risk-based food inspection manual", 2008. [Online]. Available: <http://www.fao.org/docrep/010/i0096e/i0096e00.htm>. [Accessed: 04-Dec-2011].
- [15] D. Ivnitski, I. Abdel-Hamid, P. Atanasov, and E. Wilkins, "Biosensors for detection of pathogenic bacteria," *Biosensors and Bioelectronics*, vol. 14, no. 7, pp. 599-624, 1999.
- [16] S. R. Nugen and A. J. Baeumner, "Trends and opportunities in food pathogen detection," *Analytical and Bioanalytical Chemistry*, vol. 391, no. 2, pp. 451-454, 2008.
- [17] P. Leonard et al., "Advances in biosensors for detection of pathogens in food and water," *Enzyme and Microbial Technology*, vol. 32, no. 1, pp. 3-13, 2003.
- [18] H. T. Nagle, R. Gutierrez-Osuna, and S. S. Schiffman, "The how and why of electronic noses," *IEEE Spectrum*, vol. 35, no. 9, pp. 22-31, 1998.
- [19] M. J. Pelletier, *Analytical applications of Raman spectroscopy*. Wiley-Blackwell, 1999.

- [20] G. C. Green, A.D.C. Chan, and R. A. Goubran, "An investigation into the suitability of using three electronic nose instruments for the detection and discrimination of bacteria types," in *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, pp. 1850–1853.
- [21] G. C. Green, A. D. C. Chan, and R. A. Goubran, "Dimensionality reduction of electronic nose data for bacteria discrimination," in *Proc. 30th Conference of the Canadian Medical and Biological Engineering Society*, Toronto, Canada, 2007.
- [22] G. C. Green, A. D. C. Chan, H. Dan, and M. Lin, "Using a metal oxide sensor (MOS)-based electronic nose for discrimination of bacteria based on individual colonies in suspension," *Sensors and Actuators B: Chemical*, vol. 152, no. 1, pp. 21-28, 2011.
- [23] G. C. Green, A. D. C. Chan, B. S. Luo, H. Dan, and M. Lin, "Identification of *Listeria* Species Using a Low-Cost Surface-Enhanced Raman Scattering System With Wavelet-Based Signal Processing," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 10, pp. 3713–3722, 2009.
- [24] G. C. Green, A. D. C. Chan, R. A. Goubran, B. S. Luo, and M. Lin, "A rapid and reliable method of discriminating between *Listeria* species based on Raman spectroscopy," in *Proc. IEEE Instrumentation and Measurement Technology Conference*, Victoria, BC, Canada, 2008, pp. 513–517.
- [25] G. C. Green, A. D. C. Chan, and R. A. Goubran, "Monitoring of food spoilage with electronic nose: potential applications for smart homes," in *3rd International Conference on Pervasive Computing Technologies for Healthcare*, London, UK, 2009. pp. 1–7.
- [26] G. C. Green, A. D. C. Chan, and R. A. Goubran, "Tracking Food Spoilage in the Smart Home using Odour Monitoring," in *Proc. IEEE International Symposium on Medical Measurements and Applications*, Bari, Italy, 2011, pp. 284-287.
- [27] G. C. Green, A. Chan, and R. A. Goubran, "Identification of food spoilage in the smart home based on neural and fuzzy processing of odour sensor responses," in *Proc. 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN, 2009, pp. 2625–2628.
- [28] K. S. Gracias and J. L. McKillip, "A review of conventional detection and enumeration methods for pathogenic bacteria in food," *Canadian Journal of Microbiology*, vol. 50, no. 11, pp. 883–890, 2004.
- [29] International Organization for Standardization, *Microbiology of food and animal feeding stuffs -Horizontal method for detection and enumeration of Campylobacter*, ISO 10272:2010. Geneva: ISO, 2010.
- [30] International Organization for Standardization, *Microbiology of food and animal feeding stuffs -Horizontal method for detection and enumeration of Escherichia coli O157*, ISO 16654:2001. Geneva: ISO, 2001.
- [31] International Organization for Standardization, *Microbiology of food and animal feeding stuffs -Horizontal method for detection and enumeration of Listeria monocytogenes*, ISO 11290:1998. Geneva: ISO, 1998.
- [32] J. M. Jay, M. J. Loessner, and D. A. Golden, *Modern food microbiology*. Springer, 2005.
- [33] A. N. Sharpe and A. K. Jackson, "Stomaching: a New Concept in Bacteriological Sample Preparation," *Applied Microbiology*, vol. 24, no. 2, pp. 175-178, 1972.

- [34] Health Canada, "Isolation of *Listeria monocytogenes* from all food and environmental samples," in *Compendium of Analytical Methods (Vol. 3)*, [Online]. Available: <http://www.hc-sc.gc.ca/fn-an/res-rech/analy-meth/microbio/volume3/index-eng.php>. [Accessed: 04-Dec-2011].
- [35] J. K. Struthers and R. P. Westran, *Clinical bacteriology*. Manson Publishing, 2003.
- [36] P. Singleton, *Bacteria in biology, biotechnology, and medicine*. John Wiley and Sons, 2004.
- [37] "Biomereux website." [Online]. Available: <http://www.biomerieux.com>. [Accessed: 04-Dec-2011].
- [38] V. Jasson, L. Jacxsens, P. Luning, A. Rajkovic, and M. Uyttendaele, "Alternative microbial methods: An overview and selection criteria," *Food Microbiology*, vol. 27, no. 6, pp. 710-730, 2010.
- [39] A. D. Hitchins, P. Feng, W. D. Watkins, S. R. Rippey, and L. A. Chandler, "FDA bacteriological analytical manual," *AOAC Int. Gaithersburg, USA*, pp. 10–01, 1995.
- [40] J. D. Oliver et al., "The viable but nonculturable state in bacteria," *Journal of Microbiology*, vol. 43, no. 1, pp. 93–100, 2005.
- [41] L. Sherwood, *Human Physiology: From Cells to Systems*. Brooks Cole, 2008.
- [42] E. de Boer and R. R. Beumer, "Methodology for detection and typing of foodborne microorganisms," *International Journal of Food Microbiology*, vol. 50, no. 1-2, pp. 119-130, Sep. 1999.
- [43] D. M. Kemeny and S. J. Challacombe, *ELISA and other solid phase immunoassays: theoretical and practical aspects*. John Wiley and Sons, 1988.
- [44] R. Saiki et al., "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase," *Science*, vol. 239, no. 4839, pp. 487 -491, 1988.
- [45] K. Sachse and J. Frey, *PCR detection of microbial pathogens*. Humana Press, 2003.
- [46] D. Helm, H. Labischinski, G. Schallehn, and D. Naumann, "Classification and identification of bacteria by Fourier-transform infrared spectroscopy," *Journal of General Microbiology*, vol. 137, no. 1, pp. 69-79, 1991.
- [47] D. Naumann, V. Fijala, and H. Labischinski, "The differentiation and identification of pathogenic bacteria using FT-IR and multivariate statistical analysis," *Microchimica Acta*, vol. 94, no. 1-6, pp. 373-377, 1988.
- [48] C. A. Rebuffo, J. Schmitt, M. Wenning, F. von Stetten, and S. Scherer, "Reliable and Rapid Identification of *Listeria monocytogenes* and *Listeria* Species by Artificial Neural Network-Based Fourier Transform Infrared Spectroscopy," *Applied Environmental Microbiology*, vol. 72, no. 2, pp. 994-1000, 2006.
- [49] J. Homola, "Surface plasmon resonance sensors for detection of chemical and biological species," *Chemical Reviews*, vol. 108, no. 2, pp. 462-493, 2008.
- [50] M. E. Bosch, A. J. R. Sánchez, F. S. Rojas, and C. B. Ojeda, "Recent Development in Optical Fiber Biosensors," *Sensors*, vol. 7, no. 6, pp. 797-859, Jun. 2007.
- [51] X.-L. Su and Y. Li, "A QCM immunosensor for *Salmonella* detection with simultaneous measurements of resonant frequency and motional resistance," *Biosensors and Bioelectronics*, vol. 21, no. 6, pp. 840-848, 2005.

- [52] E. Berkenpas, P. Millard, and M. Pereira da Cunha, "Detection of *Escherichia coli* O157:H7 with langasite pure shear horizontal surface acoustic wave sensors," *Biosensors and Bioelectronics*, vol. 21, no. 12, pp. 2255-2262, 2006.
- [53] S. Chemburu, E. Wilkins, and I. Abdel-Hamid, "Detection of pathogenic bacteria in food samples using highly-dispersed carbon particles," *Biosensors and Bioelectronics*, vol. 21, no. 3, pp. 491-499, 2005.
- [54] Z. Muhammad-Tahir and E. C. Alocilja, "A Disposable Biosensor for Pathogen Detection in Fresh Produce Samples," *Biosystems Engineering*, vol. 88, no. 2, pp. 145-151, Jun. 2004.
- [55] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensors and Actuators B: Chemical*, vol. 18, no. 1-3, pp. 210-211, Mar. 1994.
- [56] S. S. Schiffman and C. A. Gatlin, "Clinical Physiology of Taste and Smell," *Annual Review of Nutrition*, vol. 13, no. 1, pp. 405-436, 1993.
- [57] K. Arshak, V. Velusamy, O. Korostynska, K. Oliwa-Stasiak, and C. Adley, "Conducting polymers and their applications to biosensors: emphasizing on foodborne pathogen detection," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1942-1951, 2009.
- [58] "Smiths Detection website." [Online]. Available: <http://www.smithsdetection.com>. [Accessed: 04-Dec-2011].
- [59] "Electronic Sensor Technology website." [Online]. Available: <http://www.estcal.com>. [Accessed: 04-Dec-2011].
- [60] A. P. F. Turner and N. Magan, "Electronic noses and disease diagnostics," *Nature Reviews Microbiology*, vol. 2, no. 2, pp. 161-166, 2004.
- [61] G. S. Settles, "Sniffers: Fluid-Dynamic Sampling for Olfactory Trace Detection in Nature and Homeland Security," *Journal of Fluids Engineering*, vol. 127, no. 2, pp. 189-218, 2005.
- [62] A. K. Deisingh, D. C. Stone, and M. Thompson, "Applications of electronic noses and tongues in food analysis," *International Journal of Food Science and Technology*, vol. 39, no. 6, pp. 587-604, 2004.
- [63] T. D. Gibson et al., "Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose.," *Sensors and Actuators B: Chemical*, vol. 44, no. 1-3, pp. 413-422, 1997.
- [64] R. T. Gordon et al., "The use of canines in the detection of human cancers," *The Journal of Alternative and Complementary Medicine*, vol. 14, no. 1, pp. 61-67, 2008.
- [65] J. W. Gardner, M. Craven, C. Dow, and E. L. Hines, "The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network," *Measurement Science and Technology*, vol. 9, pp. 120-127, 1998.
- [66] M. Holmberg et al., "Bacteria classification based on feature extraction from sensor data," *Biotechnology Techniques*, vol. 12, no. 4, pp. 319-324, 1998.
- [67] A. K. Pavlou, N. Magan, D. Sharp, J. Brown, H. Barr, and A. P. F. Turner, "An intelligent rapid odour recognition model in discrimination of *Helicobacter pylori* and other gastroesophageal isolates in vitro," *Biosensors and Bioelectronics*, vol. 15, no. 7-8, pp. 333-342, 2000.
- [68] R. Dutta, E. L. Hines, J. W. Gardner, and P. Boilot, "Bacteria classification using Cyranose 320 electronic nose," *Biomedical Engineering Online*, vol. 1, pp. 4-4, 2002.

- [69] E. C. Alcocilja, N. L. Ritchie, and D. L. Grooms, "Protocol development using an electronic nose for differentiating *E. coli* strains," *IEEE Sensors Journal*, vol. 3, no. 6, pp. 801–805, 2003.
- [70] S. Balasubramanian, S. Panigrahi, C. M. Logue, M. Marchello, and J. S. Sherwood, "Identification of *Salmonella*-inoculated beef using a portable electronic nose system," *Journal of Rapid Methods and Automation in Microbiology*, vol. 13, no. 2, pp. 71–95, 2005.
- [71] J. W. Arnold and S. D. Senter, "Use of digital aroma technology and SPME GC-MS to compare volatile compounds produced by bacteria isolated from processed poultry," *Journal of the Science of Food and Agriculture*, vol. 78, no. 3, pp. 343–348, 1998.
- [72] U. Siripatrawan, "Self-organizing algorithm for classification of packaged fresh vegetable potentially contaminated with foodborne pathogens," *Sensors and Actuators B: Chemical*, vol. 128, no. 2, pp. 435–441, 2008.
- [73] S. Labreche, S. Bazzo, S. Cade, and E. Chanie, "Shelf life determination by electronic nose: application to milk," *Sensors and Actuators B: Chemical*, vol. 106, no. 1, pp. 199–206, 2005.
- [74] S. Capone, M. Epifani, F. Quaranta, P. Siciliano, A. Taurino, and L. Vasanelli, "Monitoring of rancidity of milk by means of an electronic nose and a dynamic PCA analysis," *Sensors and Actuators B: Chemical*, vol. 78, no. 1-3, pp. 174–179, 2001.
- [75] E. Smith and G. Dent, *Modern Raman Spectroscopy: A Practical Approach*. John Wiley and Sons, 2005.
- [76] A. Cao et al., "A robust method for automated background subtraction of tissue fluorescence," *Journal of Raman Spectroscopy*, vol. 38, no. 9, pp. 1199–1205, 2007.
- [77] R. A. Dalterio, M. Baek, W. H. Nelson, D. Britt, J. F. Sperry, and F. J. Purcell, "The Resonance Raman Microprobe Detection of Single Bacterial Cells From a Chromobacterial Mixture," *Applied Spectroscopy*, vol. 41, no. 2, pp. 241–244, 1987.
- [78] R. M. Jarvis and R. Goodacre, "Characterisation and identification of bacteria using SERS," *Chemical Society Reviews*, vol. 37, no. 5, pp. 931–936, 2008.
- [79] L.-P. Choo-Smith et al., "Investigating Microbial (Micro)colony Heterogeneity by Vibrational Spectroscopy," *Applied and Environmental Microbiology*, vol. 67, no. 4, pp. 1461–1469, 2001.
- [80] C. M. McEntegart, W. R. Penrose, S. Strathmann, and J. R. Stetter, "Detection and discrimination of coliform bacteria with gas sensor arrays," *Sensors and Actuators B: Chemical*, vol. 70, no. 1-3, pp. 170–176, 2000.
- [81] G. Sezonov, D. Joseleau-Petit, and R. D'Ari, "*Escherichia coli* Physiology in Luria-Bertani Broth," *Journal of Bacteriology*, pp. 8746–49, 2007.
- [82] "AlphaMOS website." [Online]. Available: <http://www.alpha-mos.com/>. [Accessed: 04-Dec-2011].
- [83] R. E. Finnigan, "Quadrupole mass spectrometers," *Analytical Chemistry*, vol. 66, no. 19, pp. 969–975, 1994.
- [84] T. C. Pearce, *Handbook of machine olfaction: electronic nose technology*. Wiley-VCH, 2003.
- [85] J. B. Chang and V. Subramanian, "Electronic noses sniff success," *IEEE Spectrum*, vol. 45, no. 3, pp. 50–56, 2008.

- [86] R. Gutierrez-Osuna, "Pattern analysis for machine olfaction: A review," *IEEE Sensors Journal*, vol. 2, no. 3, pp. 189–202, 2002.
- [87] E. Llobet, J. Brezmes, X. Vilanova, J. E. Sueiras, and X. Correig, "Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array," *Sensors and Actuators B: Chemical*, vol. 41, no. 1-3, pp. 13-21, 1997.
- [88] C. Distanto, M. Leo, P. Siciliano, and K. C. Persaud, "On the study of feature extraction methods for an electronic nose," *Sensors and Actuators B: Chemical*, vol. 87, no. 2, pp. 274-288, 2002.
- [89] J. W. Gardner, P. Boilot, and E. L. Hines, "Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach," *Sensors and Actuators B: Chemical*, vol. 106, no. 1, pp. 114-121, 2005.
- [90] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2000.
- [91] J. Ye, T. Li, T. Xiong, and R. Janardan, "Using uncorrelated discriminant analysis for tissue classification with gene expression data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, no. 4, pp. 181–190, 2004.
- [92] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic Press, 1990.
- [93] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 2003.
- [94] S. S. Schiffman, D. W. Wyrick, R. Gutierrez-Osuna, and H. T. Nagle, "Effectiveness of an Electronic Nose for Monitoring Bacterial and Fungal Growth," in *Proceedings of the 7th International Symposium On Olfaction and Electronic Nose*, Brighton, UK, 2000, pp. 1-8.
- [95] M. Trincavelli, S. Coradeschi, A. Loutfi, B. Söderquist, and P. Thunberg, "Direct identification of bacteria in blood culture samples using an electronic nose," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2884–2890, 2010.
- [96] E. Phaisangittisagul and H. T. Nagle, "Sensor selection for machine olfaction based on transient feature extraction," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 369–378, 2008.
- [97] M. S. A. Nadeem, J. D. Zucker, and B. Hanczar, "Accuracy-Rejection Curves (ARCs) for Comparing Classification Methods with a Reject Option," in *Journal of Machine Learning Research Workshop and Conference Proceedings*, vol. 8, pp. 65-81, 2010.
- [98] D. Tax and R. Duin, "Growing a multi-class classifier with a reject option," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1565–1570, 2008.
- [99] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 41–46, 1970.
- [100] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [101] C. Di Natale et al., "Electronic nose and electronic tongue integration for improved classification of clinical and food samples," *Sensors and Actuators B: Chemical*, vol. 64, no. 1-3, pp. 15–21, 2000.
- [102] A. D. C. Chan, K. B. Englehart, B. Hudgins, and D. F. Lovely, "Multiexpert automatic speech recognition using acoustic and myoelectric signals," *IEEE Transactions in Biomedical Engineering*, vol. 53, no. 4, pp. 676–685, 2006.

- [103] S. Kara, B. H. Aksebzeci, H. Kodaz, S. Günes, E. Kaya, and H. Özbilge, "Medical application of information gain-based artificial immune recognition system (IG-AIRS): Classification of microorganism species," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5168-5172, 2009.
- [104] D. O. Bayles and B. J. Wilkinson, "Osmoprotectants and cryoprotectants for *Listeria monocytogenes*," *Letters in Applied Microbiology*, vol. 30, no. 1, pp. 23-27, 2000.
- [105] Health Canada, "Listeria and Food Safety." [Online]. Available: <http://www.hc-sc.gc.ca/hl-vs/iyh-vsv/food-aliment/listeria-eng.php>. [Accessed: 04-Dec-2011].
- [106] Centers for Disease Control and Prevention, "Listeriosis: (*Listeria* Infection)." [Online]. Available: <http://www.cdc.gov/listeria/index.html>. [Accessed: 04-Dec-2011].
- [107] Canadian Food Inspection Agency, "Final Report of the Independent Investigator into the 2008 Listeriosis Outbreak." [Online]. Available: http://www.listeriosis-listeriose.investigation-enquete.gc.ca/index_e.php?s1=rpt&page=tab. [Accessed: 04-Dec-2011].
- [108] A. Champion and P. Kambhampati, "Surface-enhanced Raman scattering," *Chemical Society Reviews*, vol. 27, no. 4, pp. 241-250, 1998.
- [109] R. M. Jarvis and R. Goodacre, "Discrimination of bacteria using surface-enhanced Raman spectroscopy," *Analytical Chemistry*, vol. 76, no. 1, pp. 40-47, 2004.
- [110] B. S. Luo and M. Lin, "A portable Raman system for the identification of foodborne pathogenic bacteria," *Journal of Rapid Methods and Automation in Microbiology*, vol. 16, no. 3, pp. 238-255, 2008.
- [111] L. Robertsson and P. Wide, "Improving food quality analysis using a wavelet method for feature extraction," in *Proc. IEEE Instrumentation and Measurement Technology Conference*, Ottawa, ON, Canada, 2005, pp. 93-97.
- [112] W. Cai, L. Wang, Z. Pan, J. Zuo, C. Xu, and X. Shao, "Application of the wavelet transform method in quantitative analysis of Raman spectra," *Journal of Raman Spectroscopy*, vol. 32, no. 3, pp. 207-209, 2001.
- [113] S. G. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.
- [114] T. D. Bui and G. Chen, "Translation-invariant denoising using multiwavelets," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414-3420, 1998.
- [115] F. Ehrentreich, "Wavelet transform applications in analytical chemistry," *Analytical and Bioanalytical chemistry*, vol. 372, no. 1, pp. 115-121, 2002.
- [116] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.
- [117] Y. Liu, M. R. Smith, and R. M. Rangayyan, "The application of Efron's bootstrap methods in validating feature classification using artificial neural networks for the analysis of mammographic masses," in *Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, CA, 2004, pp. 1553-1556.
- [118] W. Zhu, X. Wang, Y. Ma, M. Rao, J. Glimm, and J. S. Kovach, "Detection of cancer-specific markers amid massive mass spectral data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, p. 14666, 2003.
- [119] L. A. Jaykus, "Challenges to developing real-time methods to detect pathogens in foods," *ASM News-American Society for Microbiology*, vol. 69, no. 7, pp. 341-347, 2003.

- [120] B. Malorny, E. Paccassoni, P. Fach, C. Bunge, A. Martin, and R. Helmuth, "Diagnostic Real-Time PCR for Detection of Salmonella in Food," *Applied Environmental Microbiology*, vol. 70, no. 12, pp. 7046-7052, 2004.
- [121] G. Dewsbury, B. Taylor, and M. Edge, "Designing safe smart home systems for vulnerable people," in *Proceedings of the Dependability and Healthcare Informatics workshop*, Edinburgh, UK, pp. 65-70, 2001.
- [122] J. F. Coughlin and J. Pope, "Innovations in health, wellness, and aging-in-place," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 4, pp. 47-52, 2008.
- [123] Statistics Canada, "Population Projections for Canada, Provinces and Territories, 2009. Catalogue No. 91-520-XIE", [Online]. Available: <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=91-520-XIE&lang=eng>. [Accessed: 04-Dec-2011].
- [124] "TAFETA website." [Online]. Available: <http://tafeta.ca/> [Accessed: 04-Dec-2011].
- [125] A. Arcelus, M. H. Jones, R. Goubran, and F. Knoefel, "Integration of smart home technologies in a health monitoring system for the elderly," in *21st International Conference on Advanced Information Networking and Applications*, Niagara Falls, ON, 2007, pp. 820-825.
- [126] A. Arcelus, C. L. Herry, R. A. Goubran, F. Knoefel, H. Sveistrup, and M. Bilodeau, "Determination of sit-to-stand transfer duration using bed and floor pressure sequences," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 10, pp. 2485-2492, 2009.
- [127] M. Vinaixa, E. Llobet, J. Brezmes, X. Vilanova, and X. Correig, "A fuzzy ARTMAP- and PLS-based MS e-nose for the qualitative and quantitative assessment of rancidity in crisps," *Sensors and Actuators B: Chemical*, vol. 106, no. 2, pp. 677-686, 2005.
- [128] R. Paolesse et al., "Detection of fungal contamination of cereal grain samples by an electronic nose," *Sensors and Actuators B: Chemical*, vol. 119, no. 2, pp. 425-430, 2006.
- [129] E. Llobet, N. El Bari, X. Correig, B. Bouchikhi, and El Barbri, "Electronic Nose Based on Metal Oxide Semiconductor Sensors as an Alternative Technique for the Spoilage Classification of Red Meat," *Sensors*, vol. 8, no. 1, pp. 142-156.
- [130] "Mathworks website." [Online]. Available: <http://www.mathworks.com/> [Accessed: 04-Dec-2011].
- [131] F. Höppner, *Fuzzy cluster analysis: methods for classification, data analysis, and image recognition*. John Wiley and Sons, 1999.
- [132] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323, 1999.
- [133] J. Stranks, *The A-Z of Food Safety*. Thorogood Publishing, 2007.
- [134] B. Iliev, M. Lindquist, L. Robertsson, and P. Wide, "A fuzzy technique for food-and water quality assessment with an electronic tongue," *Fuzzy Sets and Systems*, vol. 157, no. 9, pp. 1155-1168, 2006.
- [135] M. Broxvall, S. Coradeschi, A. Loutfi, and A. Saffiotti, "An ecological approach to odour recognition in intelligent environments," in *Proc. IEEE International Conference on Robotics and Automation*, Orlando, FA, 2006, pp. 2066-2071.
- [136] D. V. Santos, E. R. Reiter, L. J. DiNardo, and R. M. Costanzo, "Hazardous events associated with impaired olfactory function," *Archives of Otolaryngology, Head and Neck surgery*, vol. 130, no. 3, pp. 317-319, 2004.

- [137] J. Matthes, L. Groll, and H. B. Keller, "Source localization by spatially distributed electronic noses for advection and diffusion," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1711–1719, 2005.
- [138] E. Callaway et al., "Home networking with IEEE 802.15. 4: a developing standard for low-rate wireless personal area networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 70–77, 2002.
- [139] L. Pan, R. Liu, S. Peng, Y. Chai, and S. X. Yang, "An wireless electronic nose network for odours around livestock farms," in *14th International Conference on Mechatronics and Machine Vision in Practice*, Xiamen, China, 2007, pp. 211–216.
- [140] S. De Vito et al., "TinyNose: Developing a wireless e-nose platform for distributed air quality monitoring applications," in *Proc. IEEE Sensors Conference*, Lecce, Italy, 2008, pp. 701–704.
- [141] D. Hui, L. Jun-Hua, and S. Zhong-Ru, "Drift reduction of gas sensor by wavelet and principal component analysis," *Sensors and Actuators B: Chemical*, vol. 96, no. 1-2, pp. 354-363, 2003.
- [142] T. Artursson, T. Eklöv, I. Lundström, P. Maartensson, M. Sjöström, and M. Holmberg, "Drift correction for gas sensors using multivariate methods," *Journal of Chemometrics*, vol. 14, no. 5-6, pp. 711–723, 2000.
- [143] M. Holmberg, F. A. M. Davide, C. Di Natale, A. D'Amico, F. Winqvist, and I. Lundström, "Drift counteraction in odour recognition applications: Lifelong calibration method," *Sensors and Actuators, B: Chemical*, vol. 42, no. 3, pp. 185-194, 1997.
- [144] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, 2009, pp. 69-72.
- [145] M. H. Jones, R. Goubran, and F. Knoefel, "Identifying movement onset times for a bed-based pressure sensor array," in *Proc. IEEE International Workshop on Medical Measurement and Applications*, Benevento, Italy, 2006, pp. 111–114.
- [146] J. Goschnick, I. Koronczy, M. Frietsch, and I. Kiselev, "Water pollution recognition with the electronic nose KAMINA," *Sensors and Actuators B: Chemical*, vol. 106, no. 1, pp. 182-186, 2005.
- [147] M. A. Ryan, K. S. Manatt, A. V. Shevade, A. K. Kisor, L. M. Lara, and M. L. Homer, "Operation of Third Generation JPL Electronic Nose on the International Space Station," in *Proc. 39th International Conference on Environmental Systems*, Savannah, GA, 2009.
- [148] J. Goschnick, "An electronic nose for intelligent consumer products based on a gas analytical gradient microarray," *Microelectronic Engineering*, vol. 57, pp. 693–704, 2001.
- [149] J. H. Cho, Y. W. Kim, K. J. Na, and G. J. Jeon, "Wireless electronic nose system for real-time quantitative analysis of gas mixtures using micro-gas sensor array and neuro-fuzzy network," *Sensors and Actuators B: Chemical*, vol. 134, no. 1, pp. 104-111, 2008.
- [150] L. Pan, R. Liu, S. Peng, S. X. Yang, and S. Gregori, "Real-time monitoring system for odours around livestock farms," in *Proc. IEEE International Conference on Networking, Sensing and Control*, London, UK, 2007, pp. 883–888.
- [151] A. Nake, B. Dubreuil, C. Raynaud, and T. Talou, "Outdoor *in situ* monitoring of volatile emissions from wastewater treatment plants with two portable technologies of electronic noses," *Sensors and Actuators B: Chemical*, vol. 106, no. 1, pp. 36-39, 2005.

- [152] P. A. Lieberzeit and F. L. Dickert, "Chemosensors in environmental monitoring: challenges in ruggedness and selectivity," *Analytical and Bioanalytical Chemistry*, vol. 393, no. 2, pp. 467-472, 2008.
- [153] "WHO - Laboratory Biosafety Manual - Third Edition," [Online]. Available: http://www.who.int/csr/resources/publications/biosafety/WHO_CDS_CSR_LYO_2004_11/en/index.html. [Accessed: 04-Dec-2011].
- [154] C. Li, P. Heinemann, and R. Sherry, "Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection," *Sensors and Actuators B: Chemical*, vol. 125, no. 1, pp. 301-310, 2007.
- [155] M. L. Rodríguez-Méndez et al., "Fusion of three sensory modalities for the multimodal characterization of red wines," *IEEE Sensors Journal*, vol. 4, no. 3, pp. 348-354, 2004.
- [156] V. H. Tran et al., "Breath Analysis of Lung Cancer Patients Using an Electronic Nose Detection System," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1514-1518, 2010.
- [157] A. Kolk et al., "Electronic-Nose Technology Using Sputum Samples in Diagnosis of Patients with Tuberculosis," *Journal of Clinical Microbiology*, vol. 48, no. 11, pp. 4235 - 4238, 2010.