NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

Improvement of Classification Accuracy in Smell Detection Using Electronic Noses

By

Hong Chen, B.Eng., B. Med,

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Applied Science

Ottawa-Carleton Institute for
Electrical and Computer Engineering

Department of Systems and Computer Engineering

Faculty of Engineering and designing

Carleton University

Ottawa, Ontario, Canada, K1S 5B6

© Copyright

2004, HONG CHEN

Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-97402-2

Our file Notre référence

ISBN: 0-612-97402-2

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

L'auteur conserve la propriété du droit d'auteur qui protège 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.

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

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

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

Canada

PAGINATION ERROR.

TEXT COMPLETE.

ERREUR DE PAGINATION.

LE TEXTE EST COMPLET.

Abstract

This thesis investigates different signal processing and pattern recognition methods employed in the classification of smells obtained from electronic nose (Enose). While most of them perform well for simple classification tasks and large training sets, the performance degrades significantly for more complex classification tasks or under smaller training sets. For smell classification in clinical experiments, the inconvenience, high-cost and long time duration of the sampling process practically eliminates the possibility of obtaining large training sets, also the difference between the clinical smell samples is subtle, consequently the improvement of classification accuracy under such circumstances will have considerable practical importance in Enose application.

As a compromise solution to this technical challenge, a Multi-Dimension Combination (MDC) method is proposed which incorporates the classification output from each individual classifier (either feature extraction or dimension reduction method). It is shown that the MDC outperforms each individual classifier and increases the classification accuracy.

Acknowledgements

The research work reported in this thesis was carried out with the department of System and Computer Engineering at Carleton University, Ottawa, Canada, from June 2003 to August 2004. I am most grateful to co-my supervisors, Professor Rafik. A. Goubran from Carleton University and Dr. Tofy Mussivand from University of Ottawa Heart Institute during the course of the research, for their generous support of research funding, and their valuable advice and suggestions in this manuscript. I would also like to acknowledge the financial support of Teaching Assistant and Research Assistant awarded by the faculty of Graduate Studies and Research, Carleton University. I would also like to extend my appreciation to Mr. Geoff Green for his careful proofreading of the first chapter of the manuscript.

I am most indebted to my father XiXin Chen and my mother Shangxian Qin, whose education and instruction arouses my interests and potentials on academic research, and whose belief and encouragement keeps me along the path of the scientific exploration. I am equally indebted to my parents-in-law for their valuable support and priceless caring for my family.

My deepest gratitude goes to my wife Yajuan Fan, whose unselfish support and enormous sacrifice enabled me to put all my heart on the research. Throughout the years, she not only showed her great patience and understanding but also encouraged me to work for excellence and perfection.

Contents

Abstract	i
Acknowledgement	ii
Contents	iv
List of Tables	viii
List of Figures	ix
List of Abbreviations	xiii
1 Introduction	1
1.1 The Electronic nose	1
1.2 Problems statement and research objectives	3
1.3 Thesis outline	6
1.4 Summary of contributions	8
2 Overview of Electronic nose	9
2.1 Smell detection by biological nose and Enose	9
2.1.1 Smell detection by biological nose	9
2.1.2 Structure of the Electronic nose	13
2.2 Sample delivery system	14
2.3 Sensor techniques	15
2.4 Signal conditioning and processing	18
2.4.1 Baseline manipulation	18

2.4.2 Digital filter	19
2.4.3 Feature extraction	20
2.4.4 Normalization	23
2.4.5 Dimension reduction	24
2.5 Pattern Recognition	33
2.6 Wavelet feature extraction	41
2.7 Validation	44
2.8 Enose in clinical research	47
3 Experiment setup	49
3.1 Enose system	49
3.1.1 Smell samples	50
3.1.2 Cyranose 320 Enose equipment	51
3.1.3 Data processing computer	57
3.2 Experiment procedure	57
3.3 Parameter setup for Cyranose 320	58
4 Multiple classifier combining techniques	64
4.1 Introduction to MDC techniques	64
4.2 Individual classifier in MDC	69
4.3 Data combining in MDC	70
4.4 MDC in Wavelet analysis	72

5 Experiment evaluation	73
5.1 Background reference smell analysis	73
5.2 Performance of signal processing and pattern recognition methods (with adequate training samples)	76
5.2.1 Raw signal (original signal)	77
5.2.2 Baseline manipulation	79
5.2.3 Low pass filter	82
5.2.4 Normalization	85
5.2.5 Feature extraction	87
5.2.6 Dimension reduction	89
5.2.7 Pattern recognition	94
5.3 MDC with adequate training samples	94
5.4 MDC with small quantity training samples	97
5.5 MDC with variable quantity training samples	100
5.6 Simulation on sensitivity	103
5.7 Simulation on selectivity	105
5.8 Simulations on potato chips with various flavors	108
5.9 Wavelet analysis	111
6 Conclusion and future work	115
References	119

List of Tables

2.1	Enose sensor technology	16
2.2	Characteristics of some PARC algorithms	35
3.1	3.1 Fragrances samples used in the experiment	50
3.2	Various flavor of potato chips used in the experiment	51
3.3	Flow setting of Cyranose 320	59
3.4	Other settings of Cyranose 320	59

List of Figures

2.1 Human perception of odors	12
2.2 Functional structure of Enose	13
2.3 Bell-shaped window time slice	22
2.4 Radial Basis Functions	39
2.5 Radial Basis Function Network (RBF)	40
2.6 Scaling function and wavelet vector space	43
2.7 Two-stage two-band analyses Tree	44
2.8 K-fold Cross Validation	46
3.1 Smell classification system using Enose.....	49
3.2 Cyranose 320 Enose device	52
3.3 Schematic of the purge cycle	53
3.4 Schematic of the sample cycle.....	54
3.5 Sensor responses at sampling	55
3.6 Sensor responses after sampling	55
4.1 MDC using different feature extraction methods.....	68
4.2 MDC using different dimension reduction methods	69
4.3 MDC in the frequency domain with dimension reduction methods	72

5.1 Response of 32 sensors to background reference smell	74
5.2 Raw sensor response without mean centered	77
5.3 Raw sensor response with mean centered	77
5.4 Raw sensor responses (sensor 10)	78
5.5 Sensor responses with different baseline manipulation (sample1, sensor 10)...	80
5.6 Sensor responses with fractional baseline manipulation (sensor 10).....	81
5.7 Classification accuracy of baseline manipulation	82
5.8 Sensor responses to samples with SG filter (sensor 10)	83
5.9 Sensor response with different K, F values for SG Filter(Sample 2, sensor 10)	83
5.10 Classification accuracy for different K& F parameters in SG filter.....	84
5.11 Classification accuracy of filtering and local normalization methods.....	85
5.12 Different global normalization methods (Sample1, sensor 10).....	86
5.13 Classification accuracy of global normalization methods.....	87
5.14 Classification accuracy for feature extraction methods.....	89
5.15 Classification accuracy V.S PCA number (KNN method)	90
5.16 Classification accuracy for dimension reduction methods.....	91
5.17 Smellprint of different samples	92
5.18 2D PCA plot	93
5.19 3D PCA plot	93
5.20 Performance comparison between 6 feature extraction methods (FE1-FE6) and 3 MDC implementations (MDC (1) – MDC (3))	96

5.21 Performance comparison between 3 different dimension reduction methods (DR1-DR3) and 3 MDC implementations (MDC (1) – MDC (3))	97
5.22 Performance comparisons between 6 different feature extraction methods and 3 MDC implementations in small number of training samples	98
5.23 Performance comparison between 3 different dimension reduction methods and 3 MDC implementations in both 3 components (MDC3) and 2 components (MDC2) in small number of training samples	99
5.24 MDC compared with various feature extraction methods under various training sample numbers.....	101
5.25 MDC compared with various dimension reduction methods under various training sample numbers	102
5.26 Classification accuracy for various pattern recognition methods.....	103
5.27 Performance evaluations in low concentration (Dimension reduction).....	104
5.28 Performance evaluations in low concentration (Feature extraction)	105
5.29 Performance evaluations in selectivity (Dimension reduction).....	106
5.30 Performance evaluations in selectivity (Feature extraction)	107
5.31 Performance evaluation for chips' smell (feature extraction).....	108
5.32 Performance evaluation for chips' smell (dimension reduction).....	109
5.33 Raw sensor response to chip's smell (mean centered).....	110
5.34 Sensor response after baseline manipulation to chip's smell (sensor 10)	110
5.35 Frequency domain responses (1024 FFT)	111
5.36 Power Spectrum Densities	112

5.37 DWT for wavelet method.....	113
5.38 MDC performances in the frequency domain.....	113

List of Abbreviations

CHF	Congestive Heart Failure
DR	Dimension Reduction
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
ENT	Ear Nose Throat
FE	Feature Extraction
FFT	Fast Fourier Transform
FLDA	Fisher Linear Discriminant Analysis
GC	Gas Chromatograph
ICA	Independent Component Analysis
KNN	K Nearest Neighbors
LDA	Linear Discriminant Analysis
LOO	Leave One Out
LPF	Low Pass Filter
LVQ	Linear Vector Quantification
MDC	Multi-Dimension Combining
MLD	Multiple Linear Discriminant
MLP	Multilayer perceptron

MO	Metal Oxide
MOSFET	Metal-Oxide-Silicon Field Effect Transistor
MS	Mass Spectrometer
PARC	Pattern Recognition
PCA	Principle Component Analysis
PNN	Probabilistic Neural Network
QCM	Quartz Crystal Microbalance
RBF	Radial Basis Function
SAW	Surface Acoustic Wave
SG	Savitsky-Golay
SOM	Self-Organized Map
STD	Standard Variant
UTI	Urinary Tract Infections
VOC	Volatile Organic Compounds

Chapter 1

Introduction

1.1 Enose

The sense of smell (olfaction) is a primal sense for human beings as well as animals. Conventional approaches of measuring smell senses involve the use of either human odor panels to quantify and characterize the odor or some analytical instruments (such as gas chromatographs and mass spectrometers) to precisely identify the odorants producing it.

^[1] However, both approaches have their advantages and limitations, as outlined below.

Well-trained human panels can successfully discriminate smells at very low concentrations. The disadvantages of human sensory panels include subjectivity, high labor expense, poor reproducibility (i.e., results fluctuate depending on time of day, health of the panel members, prior odors analyzed, fatigue, etc.), and hazardous environment inapplicability. ^[2]

Analytical chemistry instruments such as the gas chromatograph (GC) and the mass spectrometer (MS) are also used to detect smells. GC is basically a carrier gas bottle in which the components in the gas are separated due to the stationary-phase film coated or

packed on the separation column that absorbs the different components to differing degrees.^[3] Although GC can separate the gas components with relatively high accuracy, its long testing time, high expense, relatively poor separation for several important classes and requirement for a well-equipped laboratory and well-trained staff are its main drawbacks in application.

MS separates the different gas components using a magnetic field according to the ratio of each component's mass to charge. It has almost the same advantage and disadvantages as GC, although it has a major problem in that molecules of the same mass may have quite different odors, or in other words, there is no simple smell-mass relationship for odors.^[3]

Due to the emerging requirement for qualitative, low-cost, real-time, and portable equipment to perform reliable, objective, and reproducible measures of volatile compounds and odors, a growing amount of research has been directed towards the development of instruments able to mimic the human olfactory apparatus. Such an instrument is called an electronic nose (Enose).^[2] An Enose is an instrument used for the automated detection and classification of odors, vapors and gases.^[4] In general, it consists of four components: a sample delivery system, a gas sensor array, a signal processing element, and a pattern recognition method.^[3]

Enose technology has been widely used in many applications, such as food & drink quality assessment, airport security, environment monitoring and medical diagnosis. ^[5] However, due to the non-selectivity or partially over-lapping character of the individual smell sensor, advanced signal processing and pattern recognition methods are required to improve the performance of the Enose under various circumstances.

1.2 Problem Statement and Research Objectives

The traditional smell classification process of Enose is as follows. First the smells (smell molecules) pass through the sample delivery system and excite a response in the sensor array. The response is analogue conditioned and digitized to be transferred and stored in the computer. The digitized data are then further processed by digital signal processing (DSP) and pattern recognition (PARC) methods.

The raw data collected from the sensor response are commonly contaminated with noise, including ambient interference smells, broadband intrinsic noise, and narrowband noise from electronic interference. ^[6] The raw data is also multi-dimensional (dimension according with the number of sensors), includes a large number of samples in each dimension (depending on number of training and testing samples), and has some outlier data samples (resulting from environmental change). For these factors, certain signal preprocessing techniques are usually employed to improve the performance, including: ^[6]

- Baseline manipulation
- Digital filter
- Feature extraction
- Normalization
- Dimension reduction

After preprocessing, the extracted signal features are sent into the pattern recognition (PARC) algorithm for final classification. Typical pattern recognition methods include: ^[7]

- Linear Discriminant Analysis (LDA)
- K-Nearest Neighbors (KNN)
- Multilayer Perceptron (MLP)
- Radial Basis Function (RBF)
- Probabilistic Neural Network (PNN)
- Linear Vector Quantification (LVQ)

Validation procedures are applied on each signal processing and pattern recognition method for the purpose of performance evaluation. Typical methods of validation include hold out, K-fold cross-validation and bootstrap. ^[4]

Although traditional signal processing and pattern recognition algorithms perform well for simple classification tasks or under a large number of training samples, the

performance degraded significantly when the available training samples are limited or when the classification task is considerably complex. Unfortunately, due to the high cost and long time duration of the sampling procedure in certain research areas (such as the sampling of the patient's breath in hospital for the purpose of medical diagnosis), only a small number of training samples are available. In addition, the smell difference between such clinical smell samples is quite subtle; therefore advanced signal processing and pattern recognition methods are required to successfully discriminate the smell samples.

Our approach to this classification problem is to find some relatively independent, individual classifiers with certain classification accuracy, and combine the classification results of such classifiers using the proposed Multi-Dimension Combining (MDC) techniques to improve the classification performance.

Two methods for searching qualified individual classifiers are proposed in this thesis, based on various feature extraction and dimension reduction methods respectively. For feature extraction, the following methods are tested:

- 1) Steady state: standard feature extraction method
- 2) Curve integration in both ascending and descending period
- 3) Slope in both ascending and descending period
- 4) Windowed time slicing

For dimension reduction, the following methods are employed:

- 1) Principle Component Analysis (PCA)
- 2) Independent Component Analysis (ICA)
- 3) Multiple Linear Discriminants (MLD)

In addition, three multi-classifier combination techniques are proposed, including:

- 1) Mean average
- 2) Geometric average
- 3) Squared average

The frequency feature extraction method using the wavelet method is also investigated for the possibility of being a qualified classifier for the proposed MDC techniques.

1.3 Outline of the thesis

This thesis is organized into six chapters. Chapter 1 presents a general introduction of the Enose system. It describes the technical problems and limitations of the currently used signal processing and pattern recognition methods. A summary of the thesis contribution is also included.

Chapter 2 illustrates the similarity between the human biological nose and the electronic nose. It describes the main components of the Enose system in detail, including sample delivering system, sensor array, digital signal preprocessing techniques and pattern recognition algorithms. Various verification methods are also discussed in this chapter.

Chapter 3 presents the experiment setup of the Enose system. Cyranose 320, (the Enose equipment used in experiments) is illustrated in detail while experimental smell samples and the configuration of the simulation computer are also introduced. The entire experimental procedure and parameter configuration for Cyranose 320 Enose equipment are discussed in the end.

In chapter 4, the Multi-Dimension Combining (MDC) technique is proposed. Different feature extraction and dimension reduction methods are illustrated as a candidate individual classifier for MDC methods. Various pattern recognition methods are also introduced for performance comparison purposes. Feature extraction in the frequency domain using wavelets is also introduced for comparison purposes.

Chapter 5 presents the experimental evaluation of the proposed MDC techniques, compared with the performance of each individual classifier and various other pattern recognition methods, under various training sample quantities and problem complexity. Performance of feature extraction in the frequency domain is also listed for comparison purposes.

Chapter 6 draws the conclusion and discusses future research directions.

1.4 Summary of contributions

This thesis contributes to the development of digital signal processing techniques as well as pattern recognition algorithms in the Enose system. The specific contributions include:

- 1) Evaluation of various feature extraction, dimension reduction as well as pattern recognition methods suitable for the Enose application.
- 2) Evaluation of the performance of the Enose under various smell concentration, under various sample numbers, and under various numbers of training samples.
- 3) Proposal of Multi-Dimension Combining (MDC) method. This contribution is described in “Improving the Classification Accuracy in Electronic Nose Using Multi-Dimensional Combining (MDC)” (Hong Chen, Rafik. A. Goubran, Tofy Mussivand) accepted in IEEE sensors 2004.
- 4) Verification of the developed MDC techniques through experimental tests using six different smell samples.
- 5) Compare the performance of MDC with each individual classifier as well as other pattern recognition methods, and
- 6) Investigation of feature extraction in the frequency domain and evaluation of the performance of the MDC technique in the frequency domain.

Chapter 2

Overview of Enose

This chapter describes the concept similarity between biological nose and electronic nose in section 2.1. Each part of the four components of the Enose equipment is illustrated in detail, namely sample delivery system (2.2), sensor array (2.3), signal processing (2.4) and pattern recognition (2.5). In section 2.6, the wavelet feature extraction method is illustrated. In section 2.7, various validation methods are introduced to compare the performance of each signal preprocessing and pattern recognition technique mentioned above. Finally, in section 2.8, the prospective application of Enose in the clinical field, especially the possibility of smell differentiating between CHF (congestive heart failure) patients and non-CHF subjects, is discussed.

2.1 Smell and Enose

The sense of smell is a primal sense for humans as well as animals. Smell (or olfaction) allows vertebrates and other organisms to identify food, mates, predators, and provides both sensual pleasure (the odor of flowers and perfume) as well as warnings of danger (e.g., spoiled food, chemical dangers). For both humans and animals, it is one of the most important means by which our environment communicates with us. However, compared

with seeing and hearing, smell is probably the least understood and explored of human senses, yet it is clearly important to both product and process control in many industries.^[3]

The term “Electronic nose” appeared first in 1988 and was defined as “an instrument, which comprises an array of electronic chemical sensors with partial specificity and an appropriate pattern-recognition system, capable of recognizing simple or complex odors”.

^[8] The commercial availability of the initial benchtop systems began in the early 1990’s.

Since then, a lot of research has been undertaken in sensor technology to improve the selectivity and sensitivity of the sensors to different smells, as well as in signal processing and pattern recognition analysis. ^[3] To better understand the function of the Enose, the function of the human olfactory organ, the biological nose, will be illustrated first.

2.1.1 Smell process of biological nose

The human olfactory system uses a variety of chemical sensors known as olfactory receptors, combined with automated pattern recognition incorporated in the olfactory bulb and olfactory cortex in the brain. ^[2]

No one receptor in isolation identifies a specific odor. It is the collective set of receptors combined with pattern recognition that results in the detection and identification of each

odor. Figure 2.1 illustrates the major components and function of the human olfactory system.

Figure 2.1 Human perception of odors^[10]

The major operations of olfaction can be divided into sniffing, reception, detection, recognition, and cleansing.^[2]

1) Sniffing: Sniffing brings odorant molecules from the outside world into the nose. With the aid of the turbinates (bony structures in the nose which produce turbulence), the odorant molecules are mixed into a uniform concentration and delivered to the mucus layer which also functions as a filter to remove large particles. After being dissolved in the mucus layer, they are transported to the cilia of the olfactory receptor neurons.^[9]

2) Reception: Reception involves binding the odorant molecules to the olfactory receptors. At first, the odorant molecules are temporarily bonded to the specific proteins which transport the molecules across the receptor membrane. Once across the boundary, the odorant molecules chemically stimulate the receptors. ^[9]

3) Detection & Recognition:

The chemical reaction in the receptors produces an electrical stimulus. These electrical signals are then transported by the olfactory axons through the cribriform plate (a perforated bone that separates the cranial cavity from the nasal cavity within the skull) to the olfactory bulb (a structure in the brain located just above the nasal cavity). From the olfactory bulb, the receptor response information is transmitted to the limbic system (a primitive portion of the brain that governs emotions, behavior, and memory). This gives rise to sub-conscious associations between odor and recalled memories. The olfactory information is also transmitted to the cerebral cortex (an advanced portion of the brain that processes high level information including conscious thought). This gives rise to the conscious sensation of smell which is also combined with taste to give the sensation of flavor. ^[9]

4) Cleansing: In order for the nose to respond to new odors, the olfactory receptors must be cleansed. This involves the intake of fresh air and the removal of odorant molecules from the olfactory receptors. ^[9]

2.1.2 Structure of the Enose

The electronic nose usually consists of four components: sample delivery system, sensor array and its associated electronic circuit, signal processing techniques and pattern recognition algorithms. Figure 2.2 illustrates the basic functional structure of Enose.

Figure 2.2 Functional structure of Enose

The functions of these components are in accordance with the functions of their biological counterpart: ^[2]

- a) Odorant molecules are “sniffed” by a valve and pump onto the sensor surface
- b) An electronic signal is produced upon the reaction of odorant molecules and the sensor array surface
- c) Sensor responses are further processed by signal processing techniques
- d) A classification result is reached using pattern recognition algorithms

2.2 Sample delivery system

There are three main types of sample delivery systems. ^[3] The most convenient odor delivery system is **sample-flow**, in which the sensors are placed in the vapor flow, which allows the rapid exchange of vapor and the ability to measure many samples within a short time. In the **static system**, there is no vapor flow around the sensor, and measurements are usually made on the steady-state response of the sensors exposed to vapor at a constant concentration. This system is suitable for studying the fundamental behavior of the sensor. Both the sample flow and static systems are closed units. The third method is the **direct exposure** of the sensor array to the vapor, which is an open system without sensor chamber. This system is employed for measurement in the field. The appropriate method should be selected according to the specific application. In this thesis, a static sample delivery system is employed for simplicity. ^[3]

2.3 Sensor techniques

An ideal sensor array would have each sensor highly selective and responding only to a specific chemical component. However, as the numbers of components in a smell are usually uncountable, the required number of specific sensors would be tremendous. Also, it is both expensive and difficult to build highly selective chemical sensors. An alternative approach is to use a limited number of sensors that have a broader response (non-selective) and rely on advanced information processing to discriminate between different chemicals. ^[4]

In general, Enose sensor techniques fall into five categories: piezoelectric sensors, MOSFET sensors, optical fiber sensors, spectrometry based sensors and conductivity sensors, as illustrated in table 2.1 and the following description.

- Conductivity sensors include metal oxide (MO) and polymer. Both exhibit a change in resistance when exposed to Volatile Organic Compounds (VOCs). Carbon black polymer composite sensors, which are the sensors used in this thesis, also belong to this family. ^[11]
- The piezoelectric sensor family has two members, quartz crystal microbalance (QCM) and surface acoustic wave (SAW). The principle of these sensors is that any mass change on the object (or surface of the object) will cause a decrease of the resonance frequency. Therefore, by measuring the frequency shift at the output, a tiny mass change on the object (or surface) can be calculated. ^[11]

Sensor type	Principles of operation	Sensitivity	Characteristics
Metal Oxide	Conductivity	5-500ppm	Inexpensive, microfabricated, operates at high temperature,
Conducting Polymer	Conductivity	0.1-100ppm	Operates at room temperature, very sensitive to humidity
Quartz Crystals Microbalance (QCM)	Piezoelectricity	1.0ng (mass)	Well-understood technology, interface electronics
Surface Acoustic Wave (SAW)	Piezoelectricity	1.0ng (mass)	Differential devices can be quite sensitive, interface electronics
MOSFET	Capacity charge coupling	ppm	Highly integrated, odorant reaction product must penetrate gate
OPTICAL	Fluorescence, Chemoluminescence	Low ppb	High electricity noise immunity, restricted availability of light sources
Gas Chromatography	Molecular spectrum	Low ppb	Potential analytic accuracy, sample concentration required
Mass Spectrometry	Atomic mass spectrum	Low ppb	Potential analytic accuracy, sample concentration required
Light spectrum	Transmitted light spectrum	Low ppb	Sample not consumed, requires tunable quantum-well device

Table 2.1 Enose sensor technology^[11]

- Metal-oxide-silicon field effect transistor (MOSFET) sensors are based on the principle that VOCs in contact with a catalytic metal can produce a reaction in the metal.

The reaction's products can diffuse through the gate of a MOSFET to change the electrical properties of the device. ^[11]

- Optical fiber sensors utilize glass fibers with a thin chemically active material coating on their sides or ends. A light source at a single frequency is used to interrogate the active material, which in turn responds with a change in color to the presence of VOCs to be detected or measured. ^[11]
- Spectrometry sensors include gas chromatography (GC) and mass spectrometry (MS) sensors. The quantitative measurement capability of these sensors makes them widely employed in analytical chemistry. In the Enose application, the GC column separates an odorant mixture into its molecular constituents, and the MS generates a mass spectrum for each peak. ^[11]

2.4 Signal conditioning and preprocessing

Before the sensor data can be further processed by computer for classification purposes, the following set of signal conditioning and preprocessing procedures are necessary.

- 1) The response of odor sensors (e.g. resistance change) needs to be measured and converted into electrical signal by means of interface circuits. Since interface circuits are tightly coupled to the underlying sensor technology, it is entirely application and sensor technology dependent. ^[3]

2) The electrical signal undergoes analogue signal processing to enhance its information content. The basic analogue signal processing methods include buffering, amplifying, filtering and other specific functions. ^[3]

3) The analogue signal is sampled, digitized by an A/D conversion and stored in computer memory for further digital signal preprocessing (DSP) using PC. This includes baseline manipulation, digital filter, feature extraction, normalization and dimension reduction, and will be illustrated next. ^[3]

2.4.1 Baseline manipulation

Baseline manipulation is the first stage of preprocessing and transforms the sensor response relative to its baseline response for the purpose of contrast enhancement and drift compensation. ^[4]

Assume a dataset with $e = 1 \dots N_E$ smell samples; each smell sample consists of $s = 1 \dots N_S$ sensor transients with $k = 1 \dots N_T$ time samples per sensor transient and will denoted be by $R_{e,s}(t_k)$, and the first time sample is denoted as $R_{e,s}(t_1)$

Three baseline manipulation methods are commonly employed.

a) Differential: $R_{e,s}^B(t_k) = R_{e,s}(t_k) - R_{e,s}(t_1) \quad \forall e, s, k$

b) Relative: $R_{e,s}^B(t_k) = R_{e,s}(t_k) / R_{e,s}(t_1) \quad \forall e, s, k$

c) Fractional: $R_{e,s}^B(t_k) = (R_{e,s}(t_k) - R_{e,s}(t_1)) / R_{e,s}(t_1) \quad \forall e, s, k$

As the differential method are supposed to remove the potential additive and multiplicative noise and drift

2.4.2 Digital filtering

The sensor responses are commonly embedded with various noises, therefore a digital filter can be employed to reduce noise and improve performance. ^[6] For broadband noise, Savitsky-Golay (SG) polynomial filters are usually employed. SG smoothing filters (also called digital smoothing polynomial filters or least-square smoothing filters) could be regarded as a generalized moving average filter. Normal moving average filter has no bias when the underlying function has zero second derivative (constant or linear), and will only have bias when underlying function has nonzero second derivative. ^[12] The idea of SG filter is to find filter coefficients that can preserve higher moments, or approximate the underlying function within the window as a polynomial of higher order (denoted as K here). Data values used in the moving window average is denoted as F (must be an odd number). K must be less than F , and when $K=F-1$, there is no filtering effect. ^[12]

Although SG filters are more effective at preserving the pertinent high frequency components of the signal, they are less successful than standard FIR filters at rejecting noise. ^[12] SG is the default filter employed in this thesis, various parameters (K and F) are tested for the best classification performance.

In some cases where periodic noise exists, comb notch filters are a good option to eliminate the fundamental frequency as well as its related harmonics (such as 60 Hz interference). For the known noise origin (such as temperature and humidity variations), steady state compensation or adaptive filtering could be applied, depending on the dynamic characteristics of the original noise. ^[6]

2.4.3 Feature extraction

Feature extraction is usually employed to generate descriptive parameters from the sensor's transient response and thus reduce the number of measurements per example. In most cases, this is performed by a single parameter (e.g. steady state, final and maximum response) from each sensor, disregarding the transient response. However, transient analysis has received much attention recently due to its ability to improve selectivity, reduce acquisition time, as well as increase sensor lifetime. ^[13] Transient compression methods can be grouped into three classes: ^[6]

- a) Sub-sampling methods: sample the sensor transient response (and/or derivatives) at different times during the odor exposure and/or recovery phase.
- b) Parameter-extraction methods: compress the transient response using a number of descriptors, such as rise time, recovery time, maximum/minimum responses, slope, and curve integrals.
- c) System-identification methods: fit a theoretical model (e.g. multi-exponential, auto-regressive) to the experimental transients and use the model parameters as features.

In this thesis, the following six parameter-extraction methods are used due to their simplicity and computational efficiency.

1. **Steady state** is the maximum value from the total response curve,

$$\text{defined as } R_{e,s}^{BC} = R_{e,s}^B(t_{NT}) \quad \forall e,s$$

2. **Integral 1** is the transient integral during the rising time,

$$\text{defined as } R_{e,s}^{BC} = \sum_{k=N_0}^{N_r} R_{e,s}^B(t_k) \Delta k \quad \forall e,s$$

3. **Integral 2:** is the transient integral during the falling time,

$$\text{defined as: } R_{e,s}^{BC} = \sum_{k=N_f}^{N_l} R_{e,s}^B(t_k) \Delta k \quad \forall e,s$$

4. **Slope1:** is the static response divided by the rising time,

$$\text{defined as } R_{e,s}^{BC} = R_{e,s}^B(t_{NT} - t_0) / (t_{NT} - t_0) \quad \forall e,s$$

The reason for considering the time factor is that in general, the rising time and falling time are different for each sample and therefore may contain additional information for better classification between samples.

5. **Slope 2:** is the static response divided by the falling time.

$$\text{defined as } R_{e,s}^{BC} = R_{e,s}^B(t_1 - t_{NT}) / (t_1 - t_{NT}) \quad \forall e,s$$

6. **Windowed time slice:** where the transient response is multiplied by N_w smooth, bell-shaped windowing functions and integrated with respect to time (See figure 2.3). The

idea is to capture some information about the dynamic characteristics of the response^[18].

It is defined as $R_{e,s}^{B,C} = [W^1 W^2 W^3] \quad \forall e,s$

with $W^i = \sum_{k=1}^{N_T} R_{e,s}^B(t_k) K_i(t_k) \Delta t$, and $K_i(t_k) = \frac{1}{1 + \left(\frac{t_k - c_i}{a_i} \right)^{2b_i}}$

The parameters (a_i, b_i, c_i) define the width, shape and center of the different windowing functions $k_i(t_k)$. In this thesis, $a(i)=6$; $b(i)=2$; $c(i)=0, 18, 36$ for $i=1,2,3$.

Figure 2.3 Bell-shaped window time slice

2.4.4 Normalization

Normalization can be grouped into local normalization and global normalization.

Local normalization: operates across the sensor array for each individual “sniff” in order to compensate for sample-to-sample variations caused by analyte concentration and sensor-drift, among others. The most widely used local method is vector normalization, denoted as $R_s^k = X_s^k / \sqrt{\sum_s (X_s^k)^2}$, in which the feature vector of each individual “sniff”

is divided by its norm and, as a result, is forced to lie on a hyper-sphere of unit radius. [4]

Global normalization: operates across the entire database for a single sensor; generally employed to compensate for differences in sensor scaling. Two global normalization procedures are auto-scaling and mean-centering. [4]

1) auto-scaling: in which the mean and standard variant (STD) of each feature are set to zero and unit, respectively

$$R_s^k = (X_s^k - \text{mean}(X_s)) / \text{std}(X_s)$$

2) Mean-centering: in which the range of values for each feature is set to [0, 1]

$$R_s^k = (X_s^k - \min(X_s^k)) / (\max(X_s^k) - \min(X_s^k))$$

Global methods are typically used to ensure that sensor magnitudes are comparable, preventing subsequent pattern recognition procedures from being overwhelmed by sensors with arbitrarily large values. However, both global normalization procedures can

amplify noise since all the sensors (particularly those which may not carry information) are weighted equally. ^[3]

2.4.5 Dimension reduction

The feature vector from the preprocessing stage is usually not suitable to be further processed due to its high-dimensionality and redundancy. Problems with high-dimensionality, known as “curse of the dimensionality”, imply that the number of training examples must grow exponentially with the number of features in order to learn an accurate model, which is not feasible in reality since only a limited number of examples are typically available. ^[4]

The problem of redundancy is particularly significant in Enose instruments due to the cross-selectivity of chemical gas sensors. When two or more feature dimensions are redundant, the covariance matrix of the entire dataset becomes singular and, therefore, noninvertible, which leads to numerical problems in various techniques, (e.g., quadratic classifiers and ordinary least squares). ^[4]

For these two reasons, a dimensionality reduction stage is required in most cases, using either feature extraction or feature selection. **Feature extraction** refers to a generic function of the preceding features, while **feature selection** constitutes a subset of the preceding features. ^[4]

1) Feature subset selections

Feature subset selection (FSS) is a dimensionality reduction technique that can be used to figure small sensor arrays for specific odor-measurement applications. The goal of FSS is to find an “optimal” subset of features (or sensors) that maximize information content or predictive accuracy.

2) Feature extraction

This includes Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Principle Component (ICA). Due to the simplicity and computational efficiency of feature extraction methods, these three feature extraction methods will be employed in this thesis and the details of each individual method will be further illustrated in the following.

a) PCA (Principle Component Analysis)

PCA is a multivariate statistical method expressing the response vectors in terms of linear combinations of orthogonal vectors along a new set of coordinate axes. Each orthogonal vector accounts for a certain amount of variances in the data with a decreasing degree of importance. The procedure is as follows: ^[14]

First the d-dimensional mean vector μ and d*d covariance matrix Σ are computed from

the full data set $\mathbf{X}$. ($\mu = \frac{1}{n_i} \sum_{x \in D_i} \mathbf{X}$, and $\mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T$)

Next, the eigenvectors $\mathbf{e}_i$ and eigenvalues λ_i are computed and stored according to decreasing eigenvalues. Eigenvectors $\mathbf{e}_i$ are the columns of $\mathbf{U}$ and eigenvalues λ_i are the associated value on diagonal matrix Σ . To reduce dimensionality, choose the k ($k < d$) eigenvectors with the largest eigenvalues as the representation of $\mathbf{X}$.

Finally, a $d \times k$ matrix $\mathbf{A}$ is formed, whose columns consist of the k eigenvectors. The representation of the further data by PCA consists of projecting the data onto the k -dimensional subspace according to $\mathbf{x}' = \mathbf{A}'(\mathbf{x} - \boldsymbol{\mu})$.

PCA is particularly useful in representing the high dimensional data into low dimension. Since there is often a high degree of sensor co-linearity in Enose data, the majority of the information held in response space can often be displayed using a small number of principle components (2D or 3D).^[14]

PCA is also a data dimensionality reduction technique for correlated data and can be applied to a high-dimensional dataset to explore the linear separability of the response vectors. Since the high dimensionality of the data may contain correlated information, a dimensionality reduction method such as PCA can be employed to avoid the “curse of dimension”.^[14]

PCA also has its drawbacks. First of all, PCA is a linear separation method. If the sensor output parameters are not linear, the results of PCA are not straightforward and the

interpolation of features may not be correct. Secondly, as PCA treats all sensors equally, the sensor's response may unduly influence its performance. ^[14]

b) Linear Discriminant Analysis (LDA)

As Principle Component Analysis seeks directions that are efficient for representation, Discriminant Analysis seeks directions that are efficient for discrimination. Linear Discriminant Analysis tries to yield the maximum ratio of between-class scatter to within-class scatter. Fisher Linear Discriminant is usually employed to deal with classification between two classes; for classification of more than two classes Multiple Discriminant Analysis (MDA) can be applied.

1) Fisher Linear Discriminant Analysis (FLDA) ^[14]

Suppose a set of n d -dimensional samples $\mathbf{x}_1 \dots \mathbf{x}_n$. For a linear combination of the components of $\mathbf{x}$, let $y = \mathbf{w}^t \mathbf{x}$ and a corresponding set of n samples $y_1 \dots y_n$. Geometrically, If $\|\mathbf{w}\|=1$, each y_i is the projection of corresponding $\mathbf{x}_i$ onto a line in the direction of $\mathbf{w}$.

The object of FLDA methods is to maximize inter-class difference while minimizing the within-class difference, thus improving the classification accuracy. The inter-class difference can be expressed as the difference of sample means. If $\mathbf{m}_i$ is the d -dimensional sample mean given by $\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$, then the sample mean for projected points is given

by $\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{x \in D_i} w' x = w' m_i$ and is simply the projection of $\mathbf{m}_i$. It follows that the

distance between the projected means is

$$|\tilde{m}_1 - \tilde{m}_2| = |w'(\mathbf{m}_1 - \mathbf{m}_2)|$$

For the within-class difference, rather than forming sample variances, define the scatter

for projected samples labeled ω_i by $\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2$. Thus $(1/n)(\tilde{s}_1^2 + \tilde{s}_2^2)$ is an estimate

of the variance of the data, and $\tilde{s}_1^2 + \tilde{s}_2^2$ is called the total within-class scatter of the projected samples.

The Fisher Linear Discriminant employs that linear $w' x$ for which the criterion function

$$J(w) = \frac{|\tilde{m}_1^2 - \tilde{m}_2^2|}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{w' S_B w}{w' S_w w} \text{ is maximum and independent of } \|w\|.$$

S_B is called between-class scatter matrix, defined as $S_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)'$, and S_w is

called within-class scatter matrix, defined as $S_w = S_1 + S_2$, where $S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)'$

Therefore apparently the solution is $w = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$

2) Multiple Discriminant Analysis (MDA) ^[14]

For the classification of more than two classes, the natural generalization of Fisher's Linear is Multiple Discriminant Analysis (MDA).

Assuming c-class problem ($c > 2$), MDA involves c-1 discriminant functions. Thus, the signal will be projected from a d-dimensional space to a (c-1) dimensional space, therefore tremendously reducing the vector dimensionality. Refer to the definition of S_B and S_W mentioned above; it can be proved that in MLD:

S_W can be obviously generalized to $S_W = \sum_{i=1}^c S_i$ ($c > 2$)

And S_B can be generalized to $S_B = \sum_{x \in D_i} n_i (m_i - m)(m_i - m)'$

The objective is to maximize criterion function:

$$J(w) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|w' S_B w|}{|w' S_W w|}$$

c) ICA (Independent Component Analysis)

Independent component analysis (ICA) is a multivariate statistical tool which is used to extract independent components, given only observed data that are assumed to be a linear combination of some unknown sources. The main purpose of ICA is to find a new coordinate system that makes the original signals as independent as possible, using higher order statistics from the probability densities of the data. [19]

Linear transforms can roughly be divided into two groups: second order methods and higher order methods, based on the type of statistical information used by the transform.

^[19] Second order methods use the variance within the data to estimate the transform.

When a second order method is applied, it is implicitly assumed that the data follow a normal or Gaussian distribution, since a Gaussian distribution can be completely determined by second-order statistics (i.e. mean and covariance). That is where second-order method PCA could be applied.

Higher order methods use statistical information that is not contained in the covariance structure of data, such as cumulants and moments of order greater than 2 (kurtosis, negentropy, etc). ^[19] This restriction leads to the assumption that the data is non-Gaussian, since Gaussian variables do not contain higher order information. In sum, second-order methods represent the data faithfully in the sense of reconstruction, while higher order methods give more meaningful representation.

While second-order methods provide a decorrelated representation of the data, higher order methods seek true statistical independence. Independence in general is a much stronger requirement than uncorrelatedness. Independent variables will always be uncorrelated, while the converse is usually not true, unless the variables have a joint density that is Gaussian. ^[14]

Suppose N observations $x_1, x_2 \dots x_n \dots x_N$ of M statistically independent sources $s_1, s_2 \dots s_m \dots s_M$ have been mixed together through a linear system as

$$x_n = a_{n1}s_1 + \dots + a_{nm}s_{nm} + \dots + a_{nM}s_{nM} \quad \text{For } n=1, 2 \dots N \quad (1)$$

The system of mixture is then characterized by the set of coefficients a_{nm} , under the assumption that both the mixtures x_n and sources s_m are random variables. It is convenient to use matrix notation for the observations $\mathbf{x}=[x_1, x_2 \dots x_N]^T$, and the sources $\mathbf{s}=[s_1, s_2 \dots s_M]^T$, so that the linear system in (4) becomes $\mathbf{x}=\mathbf{A}\mathbf{s}$ (2)

where $\mathbf{A}$ is a full rank $N \times M$ scalar matrix called mixing matrix. The solution is in the form $\mathbf{u}=\mathbf{W}^*\mathbf{X}=\mathbf{W}^*\mathbf{A}^*\mathbf{s} \approx \mathbf{s}$, where $\mathbf{u}$ is an estimate of $\mathbf{s}$ and $\mathbf{W}$ (called separation matrix) is an estimate of $\mathbf{A}^{-1}$. [20]

The goal is to find a matrix $\mathbf{W}$ that makes the element $\mathbf{s}$ statistically independent and that maximizes some measure of non-Gaussianity of the vectors $\mathbf{W}^*\mathbf{X}$. Non-Gaussianity may be expressed through the kurtosis (defined as the fourth moment of a random variable, equal to zero for Gaussian distribution) or the negentropy (considering that a Gaussian variable has the largest entropy with respect to all the entropies of variables of equal variances). [20]

ICA is similar to PCA but they have important differences. The directions given by PCA are mutually orthogonal and directed in the directions with the maximum variance, due to the assumptions that the density of the bivariate dataset is Gaussian. The ICA directions

do not assume a Gaussian distribution on the dataset and point in directions that provide a more meaningful interpretation of the data than PCA. ^[19]

In PCA, components are ordered according to variance contributed by each component, calculated from the relative value of its corresponding eigenvalues. This is not possible with independent components, as neither the order nor the variances of ICA can be determined. The reason is that since both $\mathbf{s}$ and $\mathbf{A}$ are unknown in (2), their relative positions in the transform are arbitrary. Therefore, a permuted version of (2), including a permutation matrix $\mathbf{P}$, is also valid $\mathbf{x} = \mathbf{AP}^{-1}\mathbf{P}\mathbf{s}$.

The application of PCA and ICA in preprocessing of Enose signals is also different.

PCA provides directions for data projection that allow an adequate decomposition on condition that the dataset follows a Gaussian distribution. However, assuming that gas-sensor array measurement data have a Gaussian joint distribution is not realistic due to various reasons. First, the complex underlying sources, such as the mechanisms of a sampling system or sensor drift, are unlikely to follow a Gaussian distribution. Assume a dataset containing repeated measurements of different types of samples, forming one cluster for each measured sample type. Even if it is possible that the measurements might follow a Gaussian joint distribution within one cluster, a Gaussian joint distribution including measurements of all clusters appears to be highly unlikely, due to the multimodal structure of data. Therefore, theoretically, PCA is not the optimal analysis tool for gas-sensor array measurement data. ICA is expected to provide more meaningful

representations of the data for non-Gaussian distributed data, which may result in improved performance. ^[19]

ICA also has its own restrictions and limitations. The first and most important requirement of ICA is that the components are assumed to be statistically independent. The second requirement is that the components are non-Gaussian distributed, which is usually the case in the real world. ^[19]

2.5 Pattern recognition algorithms

The sensor responses after signal processing are sent to pattern recognition algorithms (PARC) for classification. The simplest PARC method is called parametric techniques, including maximum-likelihood and Bayesian parameter estimation. These techniques assume that the forms of the underlying probability density functions are known (e.g. Gaussian distribution), then training data are used to optimize the target parameter of these density functions (mean and covariance matrix for example), and finally the resulting distribution will be used for classification. Despite its computational efficiency and intuitive interpretation, it is rarely the case in practice that the distribution functions can be predicted and modeled.

Linear discriminant analysis (LDA) is another parametric method used in classification. Instead of estimating the values of underlying probability density function in normal

parametric method, it assumes the known discriminant function and estimates the values of parameters of that classifier.

Nonparametric methods, on the other hand, can be employed without any assumptions of the forms of the underlying probability density function. There are two approaches for nonparametric pattern recognition. One is to estimate the densities from data samples, such as PARZEN windows method. The other is to directly estimate the posteriors probability and bypass the density estimation, such as K-nearest neighbor rule. According to Occam's razor of simplicity theory, apparently the latter will be appropriate for the classification application.

One well-known non-parametric family, called neural network, uses nonlinear function (e.g. sigmoid function) to obtain arbitrary decision regions, thus overcoming the drawback of the LDA method, which is only applicable to linearly supportable datasets (or hyperplane decision boundaries). This family includes Multilayer Perceptron (MLP), Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Linear Vector Quantization (LVQ), and Self-Organized Map (SOM), etc. Fuzzy logic is another technique which usually combines with and completes the neural network.

Pattern recognition methods can also be categorized into supervised and unsupervised. Generally, supervised methods need a training stage to build their knowledge base from the known (labeled) samples. Unsupervised methods learn to separate the different

classes routinely, without prior knowledge from training, which is closer to the way the human olfactory system works. Table 2.2 compares the characteristics of various popularly employed pattern recognition methods.

	Supervised /Not Supervised	Parametric / Not Parametric
K-means	Non-supervised	Parametric
KNN (K Nearest Neighbor)	Supervised	Non-Parametric
LDA (Linear Discriminant Analysis)	Supervised	Parametric
PCA (Principle Component Analysis)	Non-Supervised	Parametric
MLP (Multilayer Perceptron)	Supervised	Non-parametric
RBF (Radial Basis Function)	Supervised	Non-parametric
SOM (Self-Organized Map)	Non-supervised	Non-parametric
Fuzzy C-means	Non-supervised	Non-parametric

Table 2.2 Characteristics of some PARC algorithms

a) K-Nearest Neighbors (KNN)

K-nearest neighbors is a non-parametric, supervised pattern recognition method which directly estimates the posteriori probabilities $P(\omega_i | x)$ rather than estimates the probability densities first. ^[14] Assume $D^n = \{x_1, x_2, \dots, x_n\}$ denotes a set of n labeled training samples, $\{x_1', x_2' \dots x_k'\} \in D^n$ are the k nearest training samples to the test sample x , then the label of x is assigned by the majority vote of the labels from $\{x_1', x_2' \dots x_k'\}$. ^[14]

Parameter k is important in KNN methods. As k increases, the upper bound of classification accuracy gets progressively closer to the Bayes rate. In the limit as k goes to infinity, KNN becomes optimal. However, we can not allow k to increase infinitely, because we want all of the k nearest neighbors $\{x_1', x_2' \dots x_k'\}$ to be close enough to x to ensure that $P(\omega_i | x')$ is approximately the same as $P(\omega_i | x)$. Therefore k must be chosen carefully as a small fraction of the number of samples. ^[14]

b) Linear Discriminant Analysis (LDA)

Linear discriminant analysis is a non-parametric, supervised pattern recognition method. Without assuming the underlying probability distribution, it uses training samples to estimate the values of the assumed known discriminant function. ^[14]

A discriminant function is a linear combination of the components of x , denoted as $g(x) = w'x + w_0$, where w is the weight vector and w_0 is the bias or threshold weight.

For the two category case, LDA function divides the feature space by a hyperplane decision surface. The orientation of the surface is determined by the normal vector $\mathbf{w}$, and the location of the surface is determined by the bias w_0 . For generalized linear

discriminant functions, defined as $g(x) = \sum_{i=1}^{\hat{d}} a(i) y_i(x) = \mathbf{a}' \mathbf{y}$

where $\mathbf{a}$ is now a $\hat{d}$ -dimensional weight vector and the $y_i(x)$ function can be arbitrary functions of x , the result function $g(x)$ is not linear in x , but linear in y . Linear classifiers

are trained through minimization of the perceptron criterion function $J_p(\mathbf{a}) = \sum_{y \in Y} (-\mathbf{a}' \mathbf{y})$

Linear discriminant analysis is particularly useful for the linear separable samples. For samples that are not linearly separable, a nonlinear pattern recognition method such as neural network is more appropriate.

c) Multi Layer Perceptron (MLP)

Multilayer perceptron network could be regarded as the generalization of linear discriminant analysis. One main difference is that MLP uses multiple layers (one input layer, one output layer, and at least one hidden layer), while LDA uses only one single layer for weights modification. Another difference is that a nonlinear function (usually sigmoid function) is employed in MLP while not in LDA. [14]

For each layer, $a(j) = f(n(j)) = f(\sum_{i=0}^R P_i w_{ji}) = f(w'_j x)$, where P_i indexes i^{th} input, w_{ji} indexes the j^{th} weight response for i^{th} input, and $n(j)$ is the j^{th} output from the first linear processing, and $a(j)$ is the j^{th} output from the second non-linear processing. For a two layer perceptron, the overall function is $g_k(x) = f(\sum_{j=1}^{n_H} w_{kj} * f(\sum_{i=1}^d w_{ji} x_i + w_{j0}) + w_{k0})$

The simplest and most general training method employed in neural networks is the error backpropagation algorithm. This algorithm computes the error between actual output and the desired output, minimizes this error by modifying the weights of each layer backward in turn, thus bringing the actual outputs closer to the desired training samples.

It is proved that MLP with three layers is capable of solving any classification problem, ^[3] therefore this topology is adopted widely. Although the number of input nodes is typically determined by the number of sensors in the array (or more specifically, the number of extracted features), and the number of output nodes is determined by the number of odors analyzed, the number of hidden neurons (and even the number of hidden layers) are usually determined experimentally by trial and error, or by experience.

d) RBF and PNN

The radial basis function (RBF) network is the main practical alternative to the multi-layer perceptron for non-linear modeling. The transfer function for a radial neuron is:

$radbas(n) = e^{-n^2}$. See figure 2.4 for the radial basis function plot.

Figure 2.4 Radial Basis Function

From figure 2.5, the output increases as the difference between input and the weight decreases, and bias b allows the sensitivity of the function to be adjusted. The radial basis networks consist of two layers: a hidden radial basis layer of S^1 neurons, and an output linear layer of S^2 neurons.

Figure 2.5 Radial Basis Function Network (RBF)

Unlike MLP, RBF only uses standard linear techniques to minimize the criterion function. The benefit is its computational efficiency and fast convergence, but it normally needs more neurons than MLP (almost as many as input neurons). Probabilistic neural networks (PNN) is a modification of RBF networks and particularly useful for classification problems. When an input is presented, the first layer (RBF layer) computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer (compete layer) sums these contributions for each class of inputs to produce as its net output a vector of probabilities, a “compete” transfer function picks the maximum of these probabilities.

2.6 Wavelet feature extraction

In addition to the time domain feature extraction and dimension reduction methods introduced above, there is another family of feature extraction methods which analyze the signal in the frequency domain. The simplest method is to model the transient response as a sum of exponential functions. However, this method does not achieve good performance due to the fact that the exponential functions are not orthogonal.^[22] Starting from this fact, the orthogonal function of Fourier transform has been employed to model the sensor response in Enose, and the extracted features (the harmonics) are then fed into a pattern recognition classifier (often neural network) for classification.^[22] This experiment shows successful results compared with the steady state feature extraction method in the time domain. The problem of using the Fourier transform is the impossibility of localizing frequencies in the time domain, due to the fact that underlying functions $e^{j\omega t}$ are periodic and thus not localized in any particular region of the time axis. This problem can be overcome by the wavelet transform.^[23]

A basic wavelet function $\psi(t)$ is one that verifies:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z}$$

For a wavelet expansion, a finite energy signal $f(t)$ is defined by:

$$f(t) = \sum_j \sum_k a_{j,k} \psi_{j,k}(t)$$

where $a_{j,k}$ is called Discrete Wavelet Transform (DWT) of $f(t)$ and $\psi_{j,k}(t)$ is the wavelet expansion function that usually forms an orthogonal basis.

For multiresolution analysis, we define scaling function $\varphi_{j,k}(t) = 2^{j/2} \varphi_{j,k}(2^j t - k)$, whose span over k is $\nu_j = \overline{\text{Span}_k \{\varphi_{j,k}(t)\}}$ and the relation of the spanned space is illustrated in figure 2.6. Therefore, any $\varphi(t)$ in $\text{span } \nu_0$ is also in ν_1 , the space spanned by $\varphi(2t)$. This means $\varphi(t)$ can be expressed in terms of a weighted sum of shifted $\varphi(2t)$ as $\varphi(t) = \sum_n h(n) \sqrt{2} \varphi(2t - n)$, $n \in \mathbb{Z}$ which is also called Multiresolution Analysis (MRA). The coefficients $h(n)$ are a sequence of real (or complex) numbers called the scaling function coefficients (or scaling filters).

Also, we define a wavelet function which is orthogonal to the scaling function and whose span ω_j is the orthogonal complement of the scaling function space ν_j ; see the vector space spanned by both scaling and wavelet functions in figure 2.6.

Therefore, another expression of wavelet function similar to MRA is:

$$\psi(t) = \sum_n h_1(n) \sqrt{2} \varphi(2t - n) \quad n \in \mathbb{Z}, \text{ also called Multiresolution Synthesis (MRS),}$$

Figure 2.6 Scaling function and wavelet vector space

In this way, any function $g(t)$ can be expressed as:

$$g(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j, k) \psi_{j,k}(t)$$

It is important that we do not need to deal directly with scaling functions and wavelets. Only the coefficients $h(n)$ and $h_1(n)$, together with $c(k)$ and $d_j(k)$, need to be considered, which could be considered as digital filters and digital signals respectively.^[24] See figure 2.7.

$$c_j(k) = \sum_m h(m - 2k) c_{j+1}(m)$$

$$d_j(k) = \sum_m h_1(m - 2k) c_{j+1}(m)$$

Figure 2.7 Two-stage two-band analyses Tree

The purpose of wavelet feature extraction in Enose applications is to use DWT as the extracted feature for the transient sensor response. In other words, the DWT uses a set of variable window-sizes that are proportional to 2^j and operate at all dyadic scales so as to extract information about data structures that “live” at different scales. ^[23]

2.7 Validation

After signal processing and pattern recognition, the performance of each system is determined in classification accuracy, which is the ability of the system to predict the new data (different from the original training data). Validation is the procedure to evaluate the performance of various classification models and the parameter settings of each model, and select the model and parameter settings with the best performance (classification accuracy). The classification accuracy for new data is usually less than the classification accuracy in training samples, due to over-fitting ^[4].

Over-fitting is usually the result of an unreasonably large number of model parameters or excessive training iterations. The easiest way of avoiding over-fitting is called the “**Holdout**” method, which splits the available data into training and validation sets. The training set is used to learn several models with different parameter settings, and the trained model that performs best on the validation data is then selected as the final model. [4]

Although it works well in many circumstances, it has two drawbacks: First, it needs a large number of samples for both training and validation; second, the performance prediction may be misleading if we happen to get an unfortunate split. These shortcomings can be overcome by performing multiple partitions of the dataset and averaging the performance of the model across partitions. [4]

K-fold cross-validation, shown in figure 2.8, performs K data partitions in a way that each sample is eventually used for both training and validation. At each of the K splits, N/K examples are used for validation and the remaining $N*(K-1)/K$ are used for training (T. Set), where other samples are used for validation (V. Set). N is the total number of examples. When the number of folds is set equal to the number of examples ($K=N$), the method is known as **leave one out (LOO)** cross validation. [4]

V. Set	T. Set		
T. Set	V. Set	T. Set	
T. Set		V. Set	T. Set
T. Set			V. Set

Figure 2.8 K-fold cross validation (T)

Better performance estimates can be obtained with computer-intensive techniques such as **bootstrap**. Bootstrap is a statistical technique that generates multiple training-test partitions by sampling the original dataset with replacement. Examples that are not selected for training become the validation set. Besides the benefit that we can estimate the performance on small scale training samples, we can also benefit from the performance improvement by training multiple instances of a learning algorithm on different partitions and combining the results with MDC techniques, which we will introduce later. ^[4]

In this thesis, only the simplest validation method, “Hold out”, is employed to evaluate the performance of each classification method, since in general the training samples are sufficient for performance comparison. The bootstrap validation method is also under development to compare the performance of various classification methods under small number of training samples.

2.8 Enose in clinical research

Over the past few years, there have been an increasing number of attempts to diagnose diseases utilizing Enose. ^[25] Clinical researches include identification of pathogens causing infections in upper respiratory tract and ears ^[26], identification of bacteria from UTI (urinary tract infections), gastrointestinal and respiratory infections ^[25], identification of *E.coli* strains ^[27], and identification of bacteria from eye and ENT (ear, nose, throat) infections ^[28]. Other than identification of various bacteria, Enose has also been tested on lung cancer ^[29], human skin odor ^[30], and wounds ^[5].

One of the objectives of this thesis is to identify DSP preprocessing and pattern recognition approaches to differentiate the CHF (Congestive Heart Failure) patients from the non-CHF subjects by their breath smells analyzed with Enose. Congestive heart failure (CHF) is a clinical syndrome characterized by symptoms and signs of increased tissue/organ water and decreased tissue/organ perfusion. ^[31] According to the American Heart Association, CHF affects 4.7 million patients in the United States with approximately 550,000 new cases identified annually. ^[32]

The main symptoms of CHF include shortness of breath, fatigue, edema and lightheadedness. ^[33] In addition to congestive heart failure, other lung and kidney illnesses may also cause the same symptoms. The differentiation between subjects with CHF symptoms is therefore important in clinical diagnosis. Early CHF diagnostic methods will help emergency room doctors determine the real cause of the similar

symptoms and choose the appropriate remedy for each cause, as well as help limit hospitalization and keep healthcare costs down. ^[33] Therefore, an early discrimination between patients with CHF and patients with lung/kidney diseases who have the same symptoms is important in the clinic, especially in the emergency room.

However, odors in the breath are quite complicated; they can be indicative of gastrointestinal problems, lung and upper respiratory tract problems, diabetes and liver problems. ^[5] To be able to differentiate smell samples, it is essential to improve the sensitivity and selectivity of the Enose as well as adopt strict quality control of the smell samples. This thesis focuses on the first target to improve the Enose sensitivity and selectivity, especially when training samples are quite limited.

Chapter 3

Experiment setup

In this chapter, the experiment setup for smell classification using the Enose system (Cyranose 320) is illustrated. Section 3.1 describes the experimental samples and Enose equipment, section 3.2 presents the experimental procedure, and section 3.3 explains the parameter setup of Cyranose 320 and the specific considerations in the experiment.

3.1 Enose system

A smell classification system, as shown in figure 3.1, was set up for smell classification of six different common household fragrances. The entire system consists of three main parts: smell samples, Enose system (Cyranose 320 here), and data processing computer.

Figure 3.1 Smell classification system using Enose

3.1.1 Smell samples

Two sets of smell sample are experimented in this thesis. The first set is six common household fragrances, as shown in table 3.1. The second set is the four flavors of potato chips, listed in table 3.2.

Sample number	Brand	Product
#1	Brut	Splash-on lotion
#2	Colt	Toilet spray
#3	Ice	Sport after-shave
#4	Williams	Electric shave
#5	Lady Speed stick	Antiperspirant
#6	Nivea	Body moisturizer

Table 3.1 Fragrances samples used in the experiment

In the above samples, samples #1, #2, #3 and #4 are in liquid form, sample #5 is in solid form, and sample #6 is in cream form. All these smells can be differentiated by the human nose. The objective is to test the classification accuracy of the Enose system and to improve it using various signal processing and pattern recognition techniques.

Sample number	Flavors
#1	Original
#2	Sour cream and onion
#3	Light original
#4	BBQ

Table 3.2 Various flavors of potato chips used in the experiment

3.1.2 Cyranose 320 Enose equipment

Cyranose 320 is a handheld Enose instrument made by Cyrano Sciences Inc (USA) ^[10], shown in figure 3.2. It has been widely used in quality control, process measurement, hazardous materials identification, and biomedical sample discrimination, etc. It consists of four main components, namely sampling system, sensor system, signal processing techniques and pattern recognition algorithms.

Figure 3.2 Cyranose 320 Enose device^[10]

1) Sampling system

Inside Cyranose 320 a valve and pump are used to pull the vapors from either the purge inlet or the sample inlet over the sensors.^[31] Each sampling procedure can be divided into a purge cycle and a sample cycle.

a. Purge cycle

In the purge cycle, the valve is in the “open” position and allows vapors from the “purge inlet” on the side of the Cyranose 320 to be pulled across the sensor array. The vapor is then emitted at the exhaust port on the opposite side of the Cyranose 320 (figure 3.3)

This same cycle is used to refresh the sensors before and after the sample cycle. ^[31]

Figure 3.3 Schematic of the purge cycle

b. Sample cycle

In the sample cycle, the valve is switched to the sample inlet (“closed” position) and allows vapors from the “sample inlet” on the top of the Cyranose 320 to be pulled across the sensor array. The vapor is then emitted at the exhaust port on the side of the Cyranose

320 (figure 3.4). During the sample cycle, the change in resistance is measured during both transient and steady-state periods. ^[31]

Figure 3.4 Schematic of the sample cycle

2) Sensor system

Cyranose 320 consists of 32 individual thin-film carbon-black polymer composite chemiresistors (sensors) configured into an array. Each individual sensor is a composite material consisting of conductive carbon-black homogeneously blended throughout a non-conducting polymer. ^[31]

When the composite film is exposed to a vapor-phase analyte, the polymer matrix acts like a sponge and it swells up while absorbing the analyte. The increase in volume is concomitant with an increase in resistance because the conductive carbon-black pathways through the material are broken (figure 3.5).

Figure 3.5 Sensor responses at sampling ^[10]

When the analyte is removed, the polymer releases the analyte and it dries out to its original size, restoring the conductive pathways (figure 3.6). Each polymer used in the array is chemically unique and absorbs the analyte gases to a different degree, thus creating a pattern of differential responses across the sensor array.

Figure 3.6 Sensor responses after purging ^[10]

3) Signal processing techniques

a) Baseline manipulation

Baseline manipulation is the way used to manipulate the sensor response with respect to its reference analyte for the purpose of drift compensation, contrast enhancement and scaling. Cyranose 320 does not use any baseline manipulation techniques.

b) Low Pass Filter

The Savitsky-Golay filter (a low pass filter using the polynomial fit method) is implemented in Cyranose 320 to smooth the response curve and decrease high-frequency noise.

c) Feature extraction

Cyranose 320 adopted steady-state as its descriptor. Although it is computationally efficient, it discards all the transient response which might contain important information for better classification of the samples.

d) Normalization

Normalization is used to reduce the interference of sample amplitude variance. Three different normalization methods, vector normalization, auto-scaling and mean-centering, are optioned in Cyranose 320.

e) Dimension reduction

Dimension reduction is employed to reduce high dimensionality of vectors and avoid redundancy in information. Principle Component Analysis (PCA) is optioned in Cyranose 320 as the default dimension reduction method.

4) Pattern Recognition algorithms

The Cyranose 320 uses three supervised algorithms (all linear statistic methods) for building a model and predicting the unknowns, including K-means, K-nearest neighbor (KNN), and Canonical Discriminant Analysis (CDA).

3.1.3 Data processing computer

The system configuration of the desktop PC is listed below.

CPU: Pentium IV 2.8G Hz

RAM: 1G MB

OS: Window XP

Application software: PCNose (from the manufacturer)

Simulation environment: Matlab 6.5

3.2 Experiment and procedure

The whole experimental procedure is as follows:

- 1) Before sampling, the Cyranose 320 is connected to the PC via an RS-232 serial cable, and the PCNose program from the manufacturer is installed and initialized on the PC.
- 2) During sampling, the needle point of the Cyranose 320 is put into the headspace of each sample, and raw sensor responses are collected and transferred to the PC.

- 3) After each sampling, a .csv file is created on the PC containing all 32 sensor responses for the experiment sample under the preset configuration parameters.
- 4) This operation is repeated 100 times for each sample which gives a total of 600 measurements (600 csv files). All sampling process is done in random sequence to capture the maximum variability of each sample.
- 5) Upon finishing the whole sampling process, all collected data are further signal processed and classified by the developed algorithms in Matlab. Performance evaluations under each signal processing technique and pattern recognition method are recorded, analyzed and compared.
- 6) The data analysis is carried out in two categories: For the sufficient training sample case, 360 (60 for each sample) of the measurements are used for training purposes and the remaining 240 are used for performance evaluation; for the small number of training sample case, only 60 measurements are used for training purposes (10 for each sample) and the remaining 540 are used for performance evaluation.

3.3 Parameter setup of Cyranose 320

Parameter settings for Cyranose 320 in this experiment are shown in tables 3.2 and 3.3, respectively. The experiment of household fragrances is conducted under typical laboratory conditions (near-constant temperature and humidity). All samples are placed on the operation table as far away from each other as possible to avoid cross interference. Each sample is tightly sealed in its original container to assure enough headspace

concentration and only opened to the surroundings during the sampling period. Several sample considerations will be discussed later.

Procedure	Time (sec)	Pump speed
Baseline purge	5	Medium
Sample draw	10	Medium
Sample draw2	0	Medium
Snout removal	0	
1 st sample gas purge	0	High
1 st air intake purge	10	High
2 nd sample air purge	30	High
2 nd air intake purge	0	High

Table 3.3 Flow setting of Cyranose 320

Digital filtering	Off
Substrate heater	On (35 °C)
Training repeat count	1
Identify repeat count	1

Table 3.4 Other settings of Cyranose 320

1) Sample temperature

Sample temperature affects the concentration and composition of the headspace as well as the vaporization rate when sampling the headspace of each sample. Although the effect of sample temperatures on the analysis cannot be easily approximated, in a typical laboratory environment the variations of temperature can be neglected. ^[32]

2) Sample humidity

Humidity should not affect the predictive ability of C320 if it has already been properly trained, but humidity can affect the performance of the C320, since its polymer-composite sensors are non-selective. ^[32] For samples with little or no water, a silica gel cartridge can be used to condition the purge inlet. For aqueous samples, a reference sample of pure water can be employed. Due to the standard laboratory environment and various formats of the six samples, the humidity interference factor is also neglected in this experiment. ^[32]

3) Headspace enrichment

For samples with low headspace concentration, improvement can be achieved by either heating the samples or increasing the surface area of the samples. In this experiment, it is unnecessary to apply any of the above methods. ^[32]

4) Sample containers

The requirements for an eligible sample container are: ^[32]

- a) The container should be clean and not contribute to the vapor phase composition, and not contain any absorptive materials.
- b) The container should allow for enough sample value measurement to be taken without depleting the headspace during the sample process.

For convenience and simplicity, only the original package for each fragrance sample is used in the experiment.

5) Substrate temperature

The substrate temperature is typically set as low as possible, but at least 7 °C higher than the highest expected ambient operating temperature. ^[32] A temperature difference of 7 °C is required to maintain a stable substrate temperature, since the air inside the device is hotter than the ambient air. A lower substrate temperature results in stronger sensor response and also longer battery life. However, a higher substrate temperature may allow for a shorter purge time. In the experiment, the room temperature of the lab is around 25 °C, and the substrate temperature is set at 35 °C.

6) Pump speeds

C320 uses three pump speeds for sampling, approximately 50, 120 and 180 ml/min for low, medium and high settings, respectively. Pump speeds for baseline and sample

draw(s) are typically the same and initially set to “medium” when sampling headspace or “high” when sampling from an infinite source. The pump speeds are the same so that there are no pressures or other flow effects on the calculation of sensor response. A “low” setting is only used for limited headspace applications, since the headspace volume is often the limiting factor for analysis. ^[32]

Flow rate for sample purge(s) and air purge(s) are typically set to “high” so that the analyte is more quickly removed from the sensors. When set to “high” pump speed, C320 passes ambient air over the sensors and pathway so that the concentration of desorbing analyte remains low in the volume, creating a large driving force for the analyte to desorb from the sensors and pathway. ^[32] In addition, the pump on C320 is likely to create a small, negative gauge pressure in the chamber, providing an additional driving force for the analyte to desorb from the sensors. In this experiment, the default setting for pump speed is used. ^[32]

7) Time set for baseline purge, sample draw and sample purge

The default baseline draw time is 10 seconds; we use 5 seconds instead, as we find little difference in the baseline purge performance between these two settings and this hastens the process. The others are set by default, sampling draw time at 10 seconds, air intake purge at 10 seconds and sampling purge time at 30 seconds. It is observed that this time setting is adequate for most samples to reach steady state and return to baseline. ^[32]

8) Adaptive training

As the environment varies continuously in the lab (temperature, humidity, existence of other fragrances, change in the sample fragrance, etc), the sensor response will vary accordingly if the break time between sample processing is too long (longer than several hours). Adaptive training is an efficient method to improve classification accuracy under this circumstance. ^[32]

Adaptive training is a procedure to extend the longevity of the training set by replacing the oldest measurements in the model with the most recent measurements. In our experiment, we will delete the four oldest measurements at the beginning of each day from the database and replace them with the four new measurements.

Chapter 4

Multiple classifier combining techniques

In this chapter, the basic idea of multiple classifier combining techniques is introduced first in section 4.1. Two proposed individual classifier retrieving methods, based on feature extraction and dimension reduction methods respectively, are illustrated in section 4.2. Section 4.3 describes in detail the proposed multiple classifier combination techniques (Multi-Dimension Combining) employed in this thesis. Section 4.4 presents a feature extraction method in frequency domain using wavelet techniques and discusses the possibility of this method being a qualified candidate for multiple classifier combination.

4.1 Introduction to multiple classifier combination

Although different signal processing and pattern recognition methods have been tested on the classification problem, no single method appears superior to others in all cases. ^[14] Instead, these specific methods are application dependent and each sees the same dataset differently. An intuitive and natural way to improve the classification accuracy is therefore to synthesize all the available information extracted from different methods and combine them into a more robust and accurate one. Such a method is called mixture of

expert models, ensemble classifiers, or classifiers combination ^[14], and includes following categories: ^[15]

- 1) Ensemble averaging: the outputs of different predictors are linearly combined to produce an overall output;
- 2) Boosting and bagging: the same learning algorithms are used with subsets of the training examples;
- 3) Mixture of (local) experts: the outputs of different predictors are non-linearly combined through a gating network;
- 4) Ensemble on subsets of input feature: each predictor selects a group of input features.

A necessary and sufficient condition for a combined result to be more accurate than any of its individual members is that all the classifiers must be both relatively accurate and independent. Relative accuracy means that the individual classifier must have an error rate of better than random guessing on new values, while independency means that two classifiers should make different errors on new data points. ^[15]

Although it is possible to build relatively accurate classifiers, it is impractical to pretend they will be independent. Typically, they will all tend to make error decisions on samples near their classification boundary. There are three qualitative explanations on why multiple classifier combination in general works better than the best single classifier in the hypothesis space (say H). ^[15]

The first explanation is that the limited training data may not provide sufficient information for choosing the best classifier from space $\mathbf{H}$. Single classifiers based on these limited training samples are prone to misclassify on their decision boundary. The multiple classifier combination is unlikely to have similar decision boundaries for the different individual classifiers and therefore the performance will have less variance and be more robust. ^[15]

The second explanation depends on the random nature of the search in space $\mathbf{H}$. For example, finding the best weights from all the possible candidates for a neural network consistent with a set of training examples is very difficult. Instead, neural network algorithms use random initializations and local search methods. The result is normally locally minimum and sub-optimum. Different initializations and search directions in general give rise to different suboptimal sets of weights. Multiple classifier combination could be regarded as a way of compensating for this imperfect search algorithm by trying different initialization states and search directions. ^[15]

The third explanation is that if the true function is outside of $\mathbf{H}$, the multiple classifier combination may approximate it better than any individual hypothesis $\mathbf{h}$. This is due to the fact that weighted sums of hypotheses may also lie outside of $\mathbf{H}$, while each individual classifier is bound to the space $\mathbf{h}$. ^[15]

To retrieve qualified candidate classifiers, two different methods are tested. The first one creates the individual classifier from different feature extraction methods (such as steady-state, integral, slope, windowed time slice, etc) and the second one creates the individual classifier from different dimension reduction methods (such as PCA, ICA, and MLD). All the outputs from individual classifiers are then sent into MLP (Multi Layer Perceptron) for pattern recognition.

The reason for choosing different feature extraction and dimension reduction methods as individual classifiers is that each implementation of these signal processing methods will result in a certain degree of success using the same pattern recognition methods (MLP here) but none of them are totally perfect, suggesting the possible combination of these individual classifiers to reduce error occurrence. Experiment shows that the classification accuracy of each individual implementation is much higher than random guess, and further qualifies these implementations as individual classifiers.

Generally speaking, the output information that various classifiers apply can be divided into three levels: ^[16]

- 1) The abstract level: the individual classifier only outputs a unique label for each test sample
- 2) The rank level: the individual classifier ranks all the labels of each test sample in a queue with the label at the top being the first choice

3) The measurement level: the individual classifier assigns each test sample a measurement value to address the degree of confidence

Among the three levels, the measurement level contains the highest amount of information and the abstract level contains the lowest. As the Multilayer Perceptron (MLP) pattern recognition method provides the output at measurement level (i.e., the degree of confidence is the output in MLP), we choose the MLP as the universal pattern recognition method for all individual classifiers.

For combining classifiers, methods like weighted average, major voting and rank-based approaches have been already proposed.^[17] For simplicity and only for experimental purposes, three linear multiple classifier combination methods are proposed in this thesis, namely arithmetic mean average, geometric mean average and squared mean average. In spite of the simplicity of these combination methods, the improvement in performance is apparent and promising.

4.2 Individual classifier for MDC

Six features (see figure 4.1) as illustrated before have been extracted from the original dataset as candidates for the multi-classifier combination. In a similar setup, three classical dimension reduction methods (see figure 4.2) are extracted as the candidates for further multi-classifier combination.

Figure 4.1 MDC using different feature extraction methods

Figure 4.2 MDC using different dimension reduction methods

4.3 Data combining in MDC

Three linear Multi-Dimension Combination techniques are proposed in this thesis, and employed on individual classifiers retrieved from either “Feature extraction” or “dimension reduction”, with multilayer perceptron (MLP) pattern recognition and appropriate digital signal processing techniques. Our intuitive reasoning is that since various “feature extraction” and “dimension reduction” preprocessing methods illustrate the different aspects of the same dataset, and have their own degree of belief and error boundary of the same classification result, it’s likely their classification results would be independent. As a matter of fact, it’s shown in the experiment that each individual classifier has its own advantage on classification of certain classes of samples and disadvantage on the other classes of samples, compared with other classifiers. Therefore, an intelligent combination method would improve the generalization ability, tolerate the failure of an individual classifier, and increase the classification performance.

The first intuitive way is to apply an arithmetic mean average on the individual classifier

outputs, denoted as MDC1, defined as $R = \frac{1}{N} \sum_{i=1}^N c(i)$

where N: number of classifiers; c: individual classification result.

Because the individual classifier is assumed to be relatively accurate (classification accuracy above random guess) and relatively independent (does not have the exact same

error as different individual classifiers), the average mean method could reduce the error occurrence. Although the individual output from MLP could be either positive or slightly negative, it is not a problem in the arithmetic average.

The second proposed method is the use of geometric mean rather than arithmetic mean,

denoted as MDC2, defined as $R = \sqrt[N]{\prod_{i=1}^N c(i)}$, with the same parameter definition as with arithmetic average mean.

Because the classification result for an individual classifier could be positive or very slightly negative (close to zero) from the MLP output, it will cause difficulty for the root operation. When N is an odd number, the final result could be negative, which is apparently incorrect. When N is an even number, the final result could be complex, and make no sense for further comparison. Therefore the absolute value of the individual classifier is adopted in the geometric mean average method before root operation, to solve the problem.

The third average technique is proposed to cope with the negative value more efficiently than the geometric mean method. It is the squared mean value method, denoted as MDC3,

and defined as $R = \frac{1}{N} \sum_{i=1}^N c(i)^2$, with the same parameter definition as above.

In spite of the simplicity of the three proposed MDC implementations, they are found to be quite effective as they usually outperformed their individual components. More qualified individual classifiers (either “feature extraction” or “dimension reduction” methods, or feature extraction method in the frequency domain, as illustrated below) are investigated to further improve the classification accuracy of the Enose system.

4.4 MDC in wavelet analysis

In this thesis, a feature extraction method in the frequency domain using wavelet function (specifically Haar wavelet function) illustrated in chapter 2 and MDC in the frequency domain is also investigated. The raw data in the time domain are represented in the DWT (Discrete Wavelet Transform) without any signal processing, and the extracted features of DWT are sent to the various dimension reduction methods and further pattern recognition method (Multi-layer Perceptron here) for final classification, the individual classification results are combined using the proposed MDC. (See figure 4.3)

Figure 4.3 MDC in the frequency domain with dimension reduction methods

Chapter 5

Experiment Evaluation

In this chapter, the performance of the various signal processing and pattern recognition methods, including the proposed Multi-Dimension Combining (MDC) method and frequency domain feature extraction method using wavelet theory, are evaluated using the experimental dataset acquired from the sensor response of the Cyranose 320 Enose system.

Section 5.1 describes the sensor response to the background reference smell. Section 5.2 illustrates the performance of various signal processing and pattern recognition techniques. Sections 5.3 and 5.4 focus on the performance evaluation of the proposed MDC methods under adequate and small number training samples, respectively. Section 5.5 further illustrates the performance of MDC under variable training samples. Performance improvement of MDC on sensitivity and selectivity of the Enose is shown in sections 5.6.and 5.7. Finally, performance evaluation of frequency domain feature extraction using wavelets is illustrated in section 5.8.

5.1 Background reference smell analysis

Background reference smell is sampled and analyzed using the same sampling and analyzing procedure for real smell samples. The basic idea for this experiment is to verify the characteristics of the environment smell (background noise) during the sampling process for further digital processing. Figure 5.1 shows the raw data response of 32 sensors to the background reference air. Several observations can be made from this plot:

Figure 5.1 Response of all sensors to background reference smell

1) The sensor response varies according to the different sampling cycles (baseline purge, sampling draw, air intake purge and sample purge) and has the same sampling cycle as the real sell samples (see figure 5.3 and figure 5.3).

2) The sensor response has two main peaks in baseline purge and sample draw period respectively, which is different from the two main peaks in sample draw and sample purge in the real smell sample (see figure 5.2). This is because both the baseline purge and the sampling of the background smell uses the same background air.

3) The sensor response to the background reference smell is at the same level as some of the real smell samples (such as samples 2, 5 and 6 in figure 5.2 and figure 5.3) and therefore non-neglectable. The basic low pass filter can not utilize this information to improve the signal-noise-ratio (SNR). However, as it contains the characteristics of the background noise, this signal could be employed for noise cancellation or adaptive filtering.

Theoretically, if this temporary response could be deducted from the sensor response of the real smell sample, the SNR (signal to noise ratio), along with the classification accuracy, will be increased accordingly. However, the simple subtraction of the static sensor response did not increase the classification accuracy. One reason is that the background smell response is actually not a simple static process but rather a complicated stochastic process, therefore further adaptive noise canceling methods, rather than simple static deduction, may achieve better performance.

5.2 Performance of signal processing and pattern recognition methods (with adequate training samples)

In this section, all experiments are conducted under a sufficient sample number. More specifically, from the total 600 measurements, 60% (or 360 samples) are used for training and only 40% (or 240 samples) are used for testing purposes. Also, as the proposed MDC method is using MLP (Multilayer Perceptron) PARC methods, unless otherwise stated, all PARC methods used below will be MLP.

5.2.1 Raw signal (original signal)

Figure 5.2 shows the original signal response of all 32 sensors six smell samples. It is clear that the raw data are noise imbedded, amplitude varied and waveform differentiated from each other. It can be easily found that samples 1, 3, 4 are at the same amplitude level, as well as samples 2, 5 and 6. Although with only one sensor (sensor 10 here), it is difficult to differentiate all six smells (see figure 5.4), if we combine the responses of all 32 sensors, there should be adequate differences between the sample smells.

Figure 5.2 Raw sensor response of all sensors without mean centered

Figure 5.3 Raw sensor response of all sensors with mean centered

Figure 5.4 Raw sensor responses (sensor 10)

From figure 5.4, sample 2, 5 and 6 contain more high frequency noise than the other three. High frequency noise could be easily removed from the original low frequency signal by a typical low pass filter, but the high frequency signals will also be affected. Also, the sensor drift, which is not obvious in figure 5.2 but common in Enose sensors, is also in the low frequency range and therefore can not be eliminated or canceled easily by a typical low pass filter. ^[33] Thus, signal filtering is worth serious consideration and will be discussed later in section 5.2.3.

The response of the same sensor to various smell samples is different in amplitude, in rising/falling time, in response slope, and in transient response. Using any individual

feature alone, it could discriminate the samples from each other to a certain extent. An intuitive way to improve the classification accuracy is to combine multi-feature extraction methods (or multi-dimension reduction methods, introduced later) to achieve a better performance than any individual one, which is actually the proposed multi-dimension combining (MDC) methods.

5.2.2 Baseline manipulation

Three baseline manipulation methods are investigated in this experiment, namely differential, relative and fractional. Figure 5.5 shows the effect of different baseline manipulation methods of the same sensor (sensor 10) to the same sample (sample 1). As the differential method could help to reduce the differential error in the raw data, and the relative method helps to reduce the multiplicative error, the fractional method combines the advantages of the above two methods and helps to reduce both differential and multiplicative error, therefore theoretically achieving higher classification accuracy.

Figure 5.5 Sensor responses with different baseline manipulations (sample 1, sensor 10)

Figure 5.6 shows the sensor response (sensor 10) to various samples after fractional baseline manipulation (the default baseline manipulation method). Compared with figure 5.4, the main difference is that the response values after fractional baseline manipulation are mean centered on the baseline. Another difference is that the high frequency noise is somewhat reduced after the fractional baseline manipulation. This is due to the fractional baseline manipulation's capability of eliminating both the differential and multiplicative error from the original signal.

Figure 5.6 Sensor responses with fractional baseline manipulation (sensor 10)

The performance evaluation in classification accuracy for different baseline manipulation methods is revealed in figure 5.7. It is obvious that classification accuracy increases between baseline manipulation and non-baseline manipulation. Within the three baseline manipulation methods, the difference is trivial, but still the fractional baseline manipulation method outperforms the other two. Therefore, in the following analysis, unless otherwise pointed out, fractional baseline manipulation will be employed as the default baseline manipulation method.

Figure 5.7 Classification accuracy for baseline manipulation methods

5.2.3 Low Pass Filter (LPF)

As illustrated before, the function of the Savitsky-Golay low pass filter (LPF) is to remove the high frequency noise using the polynomial approximation method, while maximally preserving low frequency and transient. Figure 5.8 shows the signal after Savitsky-Golay filter. Compared with the unfiltered signal in figure 5.4, it is obvious that the high frequency noise is tremendously decreased after the filtering.

From figure 5.9, it is obviously that different filter parameters for the Savitsky-Golay filter (K& F for polynomial order and window span, respectively) will achieve different filtering output. When varying both K and F for different samples, it is found that (figure 5.10), the best classification accuracy occurred around K=1 (the green line) and F=13 ($2*6+1$) in the SG filter.

Figure 5.8 Sensor responses with SG filter (sensor 10, K=1, F=11)

Figure 5.9 Sensor response with different K, F values for SG Filter (Sample 1)

Figure 5.10 Classification accuracy for different K& F parameters in SG filter

Figure 5.11 shows the classification accuracy for different filtering and local normalization methods. From figure 5.11 it can be seen that SG filter (with parameter $K=1$ and $F=13$) performs much better than non-filtering, therefore we will stick with the Savitsky-Golay as default low pass filter (with $K=1$ and $F=13$). Normalization performance will be explained below.

Figure 5.11 Classification accuracy for filtering and local normalization methods

5.2.4 Normalization

Local normalization can be categorized as “vector normalization” and “area normalization”. The “vector normalization” process is operated vertically on each column, while “area normalization” is operated horizontally on each row. In Enose, only the “vector normalization” method is considered; the purpose is to divide each sample vector by its Euclidean norm so that it lies in a hypersphere of unit radius. From figure 5.11, vector normalization does increase the classification accuracy. However, “vector normalization” could increase performance in other pattern recognition method (such as

KNN). For performance comparison purpose, all the following analysis, if not otherwise pointed out, will adopt t vector normalization; as default local normalization method.

Global normalization can be categorized into auto-scaling and mean-centering, see figure 5.12 for the difference between the two methods.

- 1) Auto-scaling: Each example is normalized to zero mean and unit variance.
- 2) Mean-centering: Each example is normalized in the interval $[-1, 1]$.

Figure 5.12 Different global normalization methods (Sample1, sensor 10)

The classification accuracy for the global normalization method is shown in figure 5.13. It is shown that the performance difference between global normalization and non-global normalization is trivial. Also, the performance of auto-scaling is almost the same as

mean-centering. For standardization purposes, we will choose auto-scaling as the default global normalization method in the following analysis.

Figure 5.13 Classification accuracy for global normalization methods

5.2.5 Feature extraction

Feature extraction is employed to reduce the number of measurements per sample to a reasonable number. In this experiment, each sample measurement will have 101 response values per sensor, 32 sensors in all, and 600 sample measurements for one classification model with associated parameter setting. This huge number of data ($600 \times 101 \times 32$) computations is impractical and can be avoided by feature extraction and dimension reduction methods.

The most common procedure for feature extraction is to keep the steady-state or final value of each sensor response and disregard the transient, therefore reducing the data to one value for each sensor per sample. However, since the transient period may convey important information, other feature extraction methods are proposed to maximally utilize the information in the transient period.

Figure 5.14 shows the classification accuracy for different feature extraction methods. It is clear that, in general, the “windowed time slice” method will achieve higher classification accuracy compared with other feature extraction methods, due to its ability to extract more information from transient response. On the other hand, although steady-state has almost the lowest accuracy among all feature extraction methods, it is still widely used for its simplicity and computational efficiency. In the following analysis, if not otherwise pointed out, “windowed time slice” will be the default feature extraction method.

Figure 5.14 Classification accuracy for feature extraction methods

5.2.6 Dimension reduction

As illustrated in chapter 3, three dimension reduction methods are employed in this thesis. After trial and error, it is shown that when the number of features extracted for PCA is around 7-12 from the original vector dimension (96 for windowed time slicing, 32 for all others), the performance of the classifier has already reached the peak and will be up and down even when more features are added (figure 5.15). Therefore, we choose the minimum 7 as the extracted feature number. ICA will have the same number of features as PCA, as illustrated before. The number of features extracted from MLD is determined

by the number of the sample class. With 6 samples in this thesis, the final MLD number will be fixed at $6-1=5$.

Figure 5.15 Classification accuracy VS PCA number (KNN method)

The classification accuracy is shown in figure 5.16. It is obvious that under adequate training samples, the PCA method obviously outperforms the other two. Also as all three dimension reduction methods achieve high accuracy, suggesting they are prospective candidates for the further multiple dimension combination.

Figure 5.16 Classification accuracy for dimension reduction methods

After feature extraction, the extracted features are ready for the pattern recognition process. Figure 5.17 is the smellprint (bar plot) of the 32 extracted features for each sample. It is apparent that the responses of sample 1 and sample 3 are still similar to each other; the final classification will depend on the power of the pattern recognition method used.

Figure 5.17 Smellprint of different samples

Figure 5.18 and figure 5.19 are the 2D and 3D PCA plot. It can be seen that, although samples are combined in a relatively local area, there are large occurrence of overlap in both the 2D and 3D plots. It reveals that the use of only 2 or 3 vectors (extracted features) from the original 32 extracted features will not effectively separate and classify the samples.

On the other hand, with extracted features, it is not necessarily a case of the more the better. In the extreme instance, with all 32 information vector kept (no feature extraction process), the classification accuracy is also quite disappointed due to the “curse of the dimension”.

Figure 5.18 2D PCA plot

Figure 5.19 3D PCA plot

5.2.7 Pattern recognition

Instead of the MLP pattern recognition method introduced before, three other pattern recognition methods will be tested for performance evaluation purposes, namely:

- 1) KNN: K-nearest neighbor
- 2) LDA: Linear Discriminant Analysis
- 3) PNN: Probabilistic Neural Network

Multilayer Perceptron is particularly pointed out because of its convenience and advantage for measurement level multi-classifier combination. Different parameters of the neural network are tried and finally the four layer neural network construction is determined. The input layer has input neurons equal to the number of features extracted (7 in the experiment), the output layer has output neurons according to the binary representation of the number of samples to be classified (3 in the experiment), and the two hidden layers neuron numbers are determined by trial and error. In this configuration, we can achieve better classification accuracy with faster convergence.

5.3 MDC with adequate training samples

Two types of multi-dimension combining methods are tested with Enose sample data. The first one uses six different feature extraction methods and PCA for dimension reduction, the output is sent to MLP and combined for final classification decision; the second method uses windowed time slice for feature extraction and three different

dimension reduction methods for further combination, and the result is also input into MLP for final classification decision.

Feature extraction methods include:

- FE1: Steady-state
- FE2: Integration for rising time
- FE3: Time slice (3 equal segments)
- FE4: Response/time for rising time
- FE5: Response/time for falling time
- FE6: Integration for falling time

Dimension reduction methods include:

- DR1: Principle Component Analysis (PCA)
- DR2: Independent Component Analysis (ICA)
- DR3: Multiple Linear Discriminants (MLD)

Three proposed MDC implementations are:

- MDC (1): Mean average: output values are averaged for classification
- MDC (2): Geometry average: geometry means are calculated for classification
- MDC (3): Square average: squared output values are averaged for classification

For simulation under adequate training samples, the performance of different MDC methods compared with each individual feature extraction classifier is specified in figure 5.20. It's obvious that the windowed time slice achieved the highest classification accuracy, around 90%. The other feature extraction methods achieve classification accuracy from 74%-83%. It is verified that each MDC result is better than any individual classifier. In general, the classification accuracy of MDC is from 90%-93%, with the squared root average (MDC3) achieving the highest (around 93%).

Figure 5.20 Performance comparison between 6 feature extraction methods (FE1-FE6) and three MDC implementations (MDC (1) – MDC (3))

Individual dimension reduction classifiers compared with MDC outputs are listed in figure 5.21. The performance of three dimension reduction methods varies from 83% to 89%, with PCA achieves the best performance around 90%. There is an average 90% to 93% classification accuracy in MDC techniques, with squared mean average achieves the highest around 93%.

Figure 5.21 Performance comparison between 3 different dimension reduction methods (DR1-DR3) and three MDC implementations (MDC (1) – MDC (3))

5.4 MDC with small number of training samples

Considering the complex procedure and long time duration of the sampling process in the real world, normally we could not afford the luxury of hundreds of training samples.

Therefore, analysis with small number of training samples would be of more realistic importance in the real world. With the same dataset, and the same feature extraction and dimension reduction methods employed in this experiment, the only difference is that we randomly choose only 10 samples from each class for the training period (60 in total), and the remaining 90 samples from each class (540 in total) are used for verification purposes. We focus on the performance evaluation of MDC implementation compared with each individual classifier.

Figure 5.22 Performance comparisons between 6 different feature extraction methods and three MDC implementations in small number of training samples

For the small number sample experiment on the feature extraction method, the increase in classification accuracy is even more obvious. From figure 5.23, the highest classification accuracy for the single feature extraction method is still windowed time slice, around only 60%; while using MDC technique, the classification increased to around 75%, with the highest one still the squared mean average.

For the small number of sample experiment on dimension reduction methods, the increase in classification accuracy is also apparent. See figure 5.23; the naming is a little different from other plots:

Figure 5.23 Performance comparison between 3 different dimension reduction methods and 3 MDC implementations in both 3 components (MDC3) and 2 components (MDC2) in small number of training samples

It can be seen that while the first two dimension reduction methods (PCA & ICA) achieved classification accuracy around 65%, the last method (MLD) has only accuracy around 20%, which means linear discriminant analysis is not a qualified candidate for the MDC average technique. Indeed, when combining only the first two classifiers, we achieved higher classification accuracy than when combining all three classifiers (75% over 70%).

It is really a significant performance increase, considering the situation of using only very limited training samples. In the real world, the classification accuracy for the limited training samples is quite low. Although adaptive training (see chapter 3) may be applied to increase performance, the additional requirement of sample training makes it difficult to be applied in certain areas, such as clinical medicine. In clinical experiments, it is practically impossible to train the Enose system on several patients' breath samples before identifying each breath sample, so the MDC average techniques which increase performance without additional requirement for training samples are quite important to bring the Enose system to this area.

5.5 MDC with variable number of training samples

To further investigate the relations between classification accuracy and the training sample number s , experiments have been conducted with a variable number of training samples; the result is plotted on figure 5.24 and figure 5.25 for feature extraction and

dimension reduction methods respectively. It can be seen that windowed time slice performs the best among all feature extraction individual classifiers, while PCA and ICA achieve almost the same performance, which is much higher than MLD.

After MDC techniques, the overall classification accuracy, as well as the stability of the total performance is improved over each individual classifier.

Figure 5.24 MDC compared with various feature extraction methods under various training sample numbers

Figure 5.25 MDC compared with various dimension reduction methods under various training sample numbers

In conclusion, figure 5.26 evaluates the performance of various pattern recognition methods under variable number of training samples. It illustrates that the performance of MDC is much better than PNN and LDA pattern recognition methods and slightly better than KNN pattern recognition method.

It is concluded that MCD techniques using the MLP pattern recognition method can increase the overall classification accuracy to a higher or same level as the currently employed classification method, which could not be achieved by using any single individual classifier.

Figure 5.26 Classification accuracy for various pattern recognition methods

5.6 Simulation on sensitivity

Further simulation is carried out on different smell concentrations to test the ability of MDC on improving the sensitivity of the Enose. For simplicity, the simulation of low concentration smell is conducted with the needle point of the Cyranose 320 at the top of the smell container, compared with the needle point put into the container as far as possible in the high concentration smell simulation.

From figure 5.27, the MDC method (square average) outperformed all three individual classifiers (PCA, ICA and MLD), especially in small number of training samples (less than 50% or 300 samples), where an average of 5-20% increase in accuracy is obvious.

Figure 5.27 Performance evaluations in low concentration (dimension reduction)

From figure 5.28, the MDC method (square average) outperformed all six individual classifiers of various feature extraction methods. An average of 5-10% improvement in classification accuracy over the best feature extraction methods FE3 (windowed time slice) is expected. Also, the MDC method is more robust (smooth) than any of the individual methods.

Figure 5.28 Performance evaluations in low concentration (Feature extraction)

It is observed that the average classification accuracy is dramatically decreased in low concentration (about 30% from the high concentration). The reason is two-fold: First, in low concentration, the change in sensors is too small to perceive, and easier to be misjudged; second, and the most important, in low concentration the sensor response is not continuous, because the headspace of the smell is easily depleted or diluted, thus causing the final classification to be more difficult.

It is concluded that the MDC method could improve the sensitivity of Enose by combining the classification results of individual classifiers (either over feature extraction

or over dimension reduction). Even under low concentration of the smell, the improvement in classification accuracy is still obvious.

5.7 Simulation on selectivity

Another simulation is carried out on different smell sample numbers to test the ability of MDC to improve the selectivity of the Enose. Smell numbers are varied from 2 to 5, and the performance of MDC compared with individual classifiers is illustrated in figure 5.29 and figure 5.30.

Figure 5.29 Performance evaluations in selectivity (dimension reduction)

Figure 5.30 Performance evaluations in selectivity (feature extraction)

It is observed that at only two smell samples, all methods could successfully differentiate between them even at very small number of training samples (near 100%). With the number of smells increasing (from 2 to 5 in the figure), the classification accuracy decreases as well. However, MDC still outperforms the best of the individual classifiers on both feature extraction method and dimension reduction method. It is concluded that, in general, MDC can increase the selectivity of the Enose, no matter what the sample numbers or the training samples are.

5.8 Simulations on potato chips with various flavors

Another set of experiment is conducted on the smell samples from four different flavors of potato chips. The classification result is rather disappointing. None of the above digital signal processing techniques and pattern recognition methods could achieve a result much better than random guess (25% as for four smell samples). The classification result from various feature extraction and dimension reduction methods compared with MDC is illustrated below in figure 5.31 and figure 5.32. It is apparent that MDC has no advantage over individual classifiers at this circumstance, and it is within our expectation because each individual classifier does not have enough classification accuracy to combine into a more accurate one.

Figure 5.31 Performance evaluation for potato chipsc smell (feature extraction)

Figure 5.32 Performance evaluation for potato chipsc smell (dimension reduction)

Further experiments test the separability of each pair of the four flavor samples. Still the classification result is around the random guess (50% as for only two smell samples). It shows that this electronic nose device could not classify the subtle differences between any two of these four flavors, whatever the signal processing methods or pattern recognition methods we adopt.

To verify this, we trace back to the original sensor responses from the Enose device (see figure 5.33 and figure 5.34). It is shown that all sensor responses for various smell samples are identical in both amplitude and waveform, plus the sensor response is at the

same level of the background reference air, thus making it difficult for further signal processing and/or pattern recognition methods to correctly classify the smells.

Figure 5.33 Raw sensor response to potato chips smell (mean centered)

Figure 5.34 Sensor response after baseline manipulation to potato chips smell (sensor 10)

5.9 Wavelet analysis

Figure 5.35 Frequency domain responses (1024 FFT)

In addition to feature extraction in the time domain, feature extraction in the frequency domain is also tested. Although the signal frequency components are not obviously differentiated from each other in figure 5.35 (FFT transform), the power spectrum density of figure 5.36 indicates a more obvious difference.

Figure 5.36 Power Spectrogram of sensor 10 (Welch PSD Estimate)

Figure 5.37 DWT for wavelet method

Figure 5.37 illustrates the DWT of the wavelet analysis used as the compressed data in the frequency domain. In the experiment we choose the level 4 Daubechie decomposition, and use the first four DWT values as the compressed data. After the feature extraction, various dimension reduction methods (MLD, PCA and ICA), along with MLP PARC method, are employed to evaluate the performance of the MDC method.

The final result for the wavelet feature extraction method is illustrated in figure 5.38. It is clear that the performance of feature extraction in the frequency domain is not as optimal as in the time domain. An average accuracy of 50% is achieved, compared with 70-80% accuracy in the time domain.

Figure 5.38 MDC performances in the frequency domain

Among all the three dimension reduction methods (PCA, ICA and MLD), MLD achieves the best classification results at relatively high training samples. The reason is that as the signal in the time domain could be estimated in normal (Gaussian) distribution, the counterpart in the frequency domain is usually not, therefore PCA is not suitable for this case; on the other hand, the assumption of the independence is also not guaranteed, therefore ICA is not appropriate under this circumstance. MLD, however, only assumes the linear separation in the data and in this case could give the best approximation.

At relatively small training samples, PCA and ICA performs much better than MLD. When MDC is applied, it could achieve the best performance in both small and large training samples, therefore combining the advantages of all individual classifiers and avoiding their shortcoming.

Chapter 6

Conclusions and future work

This thesis investigated the ability of the Multi-Dimension Combining (MDC) method to improve the classification accuracy of the Enose system, and compared the performance of MDC with the performance of each individual classifier—either individual feature extraction or individual dimension reduction method. It also compared the performance of MDC with other pattern recognition methods such as KNN, LDA and PNN.

As a primary experimental stage for further application of the Enose system in clinical experiments, the performance of the Enose system with small number of training samples was particularly investigated. In clinical circumstances, smell samples (such as breath samples) from subjects (either patients or controls) would be under strict restriction and very costly and inconvenient to acquire. Therefore, performance under a large number of training samples would be of theoretical importance only and performance under a small number of training samples is of practical importance in this application.

Although individual classifiers could achieve satisfying classification results under adequate training samples (in average >90%), they fail to offer the same satisfying classification results under small number of training samples (around 65% at the best).

By applying the proposed MDC average techniques on the same dataset, the average classification accuracy has been improved in both adequate training samples and small number of training samples. Under adequate training samples, the best classification accuracy of an individual classifier is around 90%, compared with the highest classification accuracy of MDC, which is near 95%, or a 5% increase. When training sample numbers are small, the improvement is more apparent. While the best classification accuracy of an individual classifier is around 65%, the highest classification accuracy of MDC is near 75%, or a 10% increase. Considering it uses only a small number of training data, and the classification accuracy improvement is quite steady when training samples increase, it is really an exciting improvement, which makes it possible for the Enose to be tested in clinical experiment.

Compared with other PARC methods such as KNN, LDA and PNN, MDC also shows its advantage. When total training samples are over 100 (or 16 for each smell), MDC always outperforms PNN and LDA (in general about 5%-20% increase in classification accuracy). Compared with KNN, it still has a 5%-10% increase in classification accuracy at most of the time, while at certain points the performances are similar or even slightly lower.

Experiments were also conducted on various smell concentrations to test the sensitivity improvement of the Enose, and on various smell sample numbers to test the selectivity improvement in the Enose. In all occasions, the MDC method consistently outperforms

the best of the individual classifiers. Further experiments on more subtle smell differences of four flavors potato chips illustrate that when sensor response is at the same level of the background reference air, or when the smell concentration is not adequate, sensors could make a lot of errors at classification. Therefore enough smell concentration is important in successful classification of various smells.

Besides analysis in the time domain, feature extraction method in the frequency domain was also investigated using the wavelet method. The idea is to use DWT of the wavelet function to represent the typical characteristic of the transient signal. Since a feature extracted from the frequency domain is totally different from the time domain, it may be considered as an individual classifier candidate for the proposed MDC. Although the experimental result is not as optimum as that in the time domain, the performance improvement of MDC over each individual component classifier is still obvious.

It is worth noticing that the original classification form Cyranose 320 is only qualitative, not quantitative, therefore the result between proposed MDC and Cyranose 320 can not be compared directly. However, by the observations in the experiment result, in general proposed MDC are more accurate and robust than the result from the Cyranose.

Future work to improve the performance of the Enose system could rely on two developments. First is the optimization and improvement of the existing signal processing and pattern recognition methods, to maximally utilize the information, as well as decrease

the noise interference. Fuzzy logic, genetic algorithm and adaptive noise canceling are among the several candidates for further research.

The other solution is to employ the proposed MDC methods more effectively and more efficiently. One suggested improvement is to use nonlinear MDC methods rather than the linear ones proposed here. Another possible way is to combine the information collected from the “electronic tongue” and “electronic nose” using the proposed MDC methods. “electronic tongue” is the equipment for liquid sample discrimination based on pulse voltammetry.^[34] Mimicking the human’s capability of synthesizing the signal from both nose and tongue, it is expected that the fusion of “Enose” and “Etongue” using various MDC methods will achieve better performance than any individual one. ^{[35][36]}

REFERENCES

- [1] Ranjit Singh, "An intelligent system for odor discrimination," Proceedings of the first IEEE International Workshop on Electronic Design, Test and Application (Delta'02), pp. 1-3, 2002.
- [2] Paul E. Keller, "Physiologically inspired pattern recognition for electronic noses," SPIE Applications and Science of Computational Intelligence II, Orlando, FL, USA, 5-9 April 1999.
- [3] T.C.Pearce, S.S.Schiffman, H.T.Nagle, J.W.Garden, *Handbook of machine olfaction—Electronic Nose Technology*, WILEY-VCH, 2003.
- [4] Richard Gutierrez-Osuna, "Pattern analysis for machine olfaction: A reviews," IEEE Sensors Journal, vol.2, no.3, pp. 189-202, June 2002.
- [5] Paul E. Keller, "Electronic noses and their applications," IEEE Technical Applications Conference and Workshops Northcon95, pp. 116-119, Oct. 1995
- [6] A.Perera, T. Sundic, A.Aprdo, R.Gutierrez-Osuna, "A portable electronic nose based on embedded PC technology and GNU/Linux: Hardware, software and applications," IEEE Sensors Journal, vol. 2, no. 3, pp. 235-245, June 2002.
- [7] Vladimir Cherkassky & Filip Mulier, *Learning from data—Concepts, Theory, and Methods*, John Wiley & sons, inc. 1998.
- [8] J.Brezmes, N.Canyellas, E.Llobet, X.Vilanova, X.Correig, "Application of artificial neural networks to the design and implementation of electronic olfactory systems," NEUREL 2000. Proceedings of the 5th Seminar on, pp. 75 – 80, September 2000.
- [9] P.E. Keller, "Overview of electronic nose algorithms," IJCNN '99, vol. 1, pp. 309-310, 1999.
- [10] <http://cyranosciences.com/applications/medical.html>
- [11] Nagle, H.T.; Gutierrez-Osuna, R.; Schiffman, S.S, "The how and why of electronic noses," IEEE Spectrum, vol. 35, issue. 9, pp. 22-31, September, 1998.
- [12] Orfanidis, S.J, *Introduction to Signal Processing*, Prentice-Hall, Eaglewood Cliffs, NJ, 1996.

- [13] Graham E. Searle, Julian W. Gardner; "System identification of electronic nose data from Cyanobacteria experiment," *IEEE Sensors Journal*, vol. 2, no. 3, pp. 219-229, June 2002.
- [14] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern classification*, Wiley Interscience, 2001.
- [15] Matteo Pardo and Giorgio sberveglieri, "Learning form data: A tutorial with emphasis on modern pattern recognition methods," *IEEE Sensors Journal*, vol. 2, no. 3, pp. 203-217, June 2000.
- [16] Lei Xu, Adam Krzyzak, Ching Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE transaction on systems, man and cybernetics*, vol.22, no.3, pp. 418-435, MAY/JUNE 1992.
- [17] Jiun-Hung Chen and Chu-Song Chen, "Reducing SVM classification time using multiple mirror classifiers," *IEEE transactions on systems, man and cybernetics*, vol. 34, no. 2, pp. 1173-1183, April 2004.
- [18] Ricardo Gutierrez-Osuna and H. Troy Nagle, "A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors," *IEEE transactions on systems, man and cybernetics-Part B: Cybernetics*, vol.29, no.5, pp. 626-632, October 1999.
- [19] Martin Kermit and Oliver Tomic, "Independent component analysis applied on gas sensor array measurement data," *IEEE Sensor Journal*, vol. 3, issue 2, pp. 218-228, April 2003.
- [20] E.Martineili; C. Falconi; A. D'Amico; C.Di Natale, "Preprocessing of Electronic Nose data by independent component analysis," *IEEE Sensors Journal*, Proceedings of IEEE, vol. 2, pp. 1339-1342, June 2002.
- [21] <http://www.mathworks.com/>
- [22] Cosimo Sistante and Pietro Siciliano, "A comparison between feature extraction methods of an electronic nose responses," *ICECS 2001. The 8th IEEE International Conference on September 2001* vol. 3, pp. 1243-1246.
- [23] Cosimo Sistante and Pietro Siciliano, "On the study of feature extraction methods for an electronic nose," *Sensors and Actuators B* 87 pp. 274-288, 2002.
- [24] C. Sydney Burrus, Ramesh A. Gopinath, and Haitao Guo, *Introduction to Wavelets and Wavelet transforms-A primer*, Prentice Hall, 1997.

- [25] V.S.Kodogiannis, P.Chountas, A. Pavlou, I.Petrounias, etc, "Electronic nose: Clinical diagnosis based on soft computing methodologies," 2002 First International IEEE Symposium "Intelligent Systems", pp.254-259, September 2002.
- [26] Julian W. Gardner, Hyun Woo Shin, Evor L. Hines, "An electronic nose systems to diagnose illness," *Sensors and Actuators B* 70, pp. 19-24, 2000.
- [27] Evangelyn C. Alocilja, Nichole L. Ritchie, and Daniel L. Grooms, "Protocol development using an electronic nose for differentiating *E.Coli* strains," *IEEE Sensors Journal*, vol. 3, no. 6, pp.801-805, December 2003.
- [28] Pascal Boilot, Evor L. Hines, Julian W. Gardner, etc, "Classification of bacteria responsible for ENT and eye infections using the Cyranose systems," *IEEE Sensors Journal*, vol. 2, no. 3, pp. 247-253, June 2002.
- [29] Corrado Di Natale, Antonella Macagnano, Eugenio Martinelli, etc, "Lung cancer identification by analysis of breath by means of an array of non-selective gas sensors," *Biosensors and Bioelectronics* 18, pp. 1209-1210, 2003.
- [30] Corrado Di Natale, Antonella Macagnano, Eugenio Martinelli, etc, "Human skin odor analysis by means of an electronic nose," *Sensor and Actuators B* 65, pp. 216-219, 2000.
- [31] User manual of the Cyranose 320 Electronic nose
Edition 4, Revision 5
- [32] The practical Guide to the Cyranose 320
Part number 02-40152, Revision C
- [33] E. Schaller; J.O.Bosset and F. Escher, "Instability of conducting polymer sensors in an electronic nose system," *Analisis*, vol. 28, pp. 217-227, 2000.
- [34] Peter Wide, Frederik Winquist, Pontus Bergsten, Emil M Petriu, "The human-based multi-sensor fusion method for artificial nose and tongue sensor data," *IEEE Instrument and Measurement Technology Conference*, St. Paul, Minnesota, USA, May 1998.
- [35] T. Sundic, S. Marco, J Samitier, P.Wide, "Electronic tongue and electronic nose fusion in classification with neural networks and fuzzy logic based models," *IMTC 2000, Proceedings of the 17th IEEE*, vol. 3, pp. 1474- 1479, May 2000.

- [36] Peter Hauptmann ; Ralf Borngraeber; Jens Schroeder, "Artificial electronic tongue in comparison to the electronic nose –state of the art and trend," 2002 IEEE/EIA International Frequency Control Symposium and Exhibition, pp. 22-29, 2002.