Thesis access form

Breath analysis: Methodology towards a fieldable breath analysis device

by

Gavin James Blackburn

A Doctoral Thesis

Submitted in partial fulfillment of the requirements
for the award of
Doctor of Philosophy of Loughborough University

Certificate of originality

Abstract

In this work lung cancer is introduced along with the current detection methods. The inadequacies of the current situation are highlighted along with the need for better detection technologies that would allow for a more rigorous testing regime to be implemented. Metabolism and metabolites are introduced as potential biomarkers. The advanced detection techniques mass spectrometry (MS) and differential mobility spectrometry (DMS) are introduced and discussed with regard to being a fieldable device. The methods applicable to processing data generated by these instruments are discussed. Finally the research objectives are highlighted.

The science of breath sampling is discussed along with the considerations when engaging in breath analysis research. Sampling and trapping of volatile organic compounds (VOCs) is discussed with particular emphasis on the adaptive breath sampler which was used in this work. The benefits of a dual detector instrument allowing for analysis of a single sample using both MS and DMS are outlined.

The design and implementation of a parallel, two detector system is outlined including the intricacies of balancing the two columns that operate at different pressures and developing a mount Processing DMS data currently lags behind the current hardware available as there are no methods that allow the full data surface to be utilised. This work outlines a method for transforming DMS data from three dimensions to two dimensions while retaining the full information contained within the data surface. This method was tested with generated data sets to show its' utility and compared to the current standard processing method using real data sets.

An understanding of all aspects of a clinical research project is vital to ensure the smooth running and completion of the project. The currently required documentation for an outside researcher to work within the NHS are detailed along with the expected timeframe for each step of designing, gaining ethical approval and implementing the research. The use of Gantt charts and work flow diagrams is highlighted and examples are given.

An initial inspection of the data produced by a pilot study shows that there a several challenges that must be overcome, these are contamination and artefact peaks, retention time shifting, unresolved peaks, differing intensities in similar samples and the complexities of correctly identifying compounds found in breath samples. These are discussed and a workflow is highlighted.

Dedication

This thesis is dedicated to my wife, Sarah. Without her love, support and encouragement it would not have been possible.

Acknowledgements

Firstly I would like to acknowledge my supervisor Paul Thomas. While his help and guidance with this work is the most evident thing he has taught me how to be truly inquisitive and engaged in research and for that I am most grateful.

A multidisciplinary research programme requires many different skills and I would like to acknowledge the following people for their help;

Stephen Fowler for his guidance in the clinical aspects of this work

Cristina Guallar-Hoyas for her help with collecting and analysing the breath samples

Craig MacKay for his help collecting samples

Victor Bocos-Bintintan and Victor Moll for producing the real data sets for testing the Euclidian transform processing method

Sultan Shair for collaborating on a program to automate the Euclidian transform method

James Reynolds for his help with the instrumentation

I would also like to thank my industrial supervisor Ian Wilson and Paul Holme for their help and guidance throughout this work.

Last, but certainly not least, I would like to thank my parents, Steve and Yvonne Blackburn. They have been there to support me, not just throughout this work but throughout my whole life and for that I am eternally grateful.

Table of Contents

Table of Contents 7

1 An introduction to lung cancer, current detection methods and applicable detection technologies for fieldable devices 15

1.1 Lung cancer 15

1.1.1 Types of lung cancer 16

1.1.1.1 Adenocarcinoma 16

1.1.1.2 Squamous cell carcinoma 18

1.1.1.3 Large-cell carcinoma 19

1.1.1.4 Small-cell carcinoma (oat cell carcinoma) 19

1.1.1.5 Mesothelioma 21

1.1.2 Causes of lung cancer 22

1.1.3 Prevalence of lung cancer 23

1.2 Current lung cancer diagnosis techniques 26

1.2.1 Spirometry 27

1.2.2 Chest X-ray (CXR) and X-ray computed tomography (CT) scan 27

1.2.3 Bronchoscopy 28

1.2.4 Biopsy 29

1.2.5 The need for better diagnostic techniques 30

1.3 Biochemical basis of cancer targets 31

1.3.1 Metabolism and metabolites 33

1.3.1.1 Catabolism, catabolites and reactive oxygen species (ROS) 33

1.3.1.2 Anabolism and anabolites 35

1.4 Volatile organic compounds and lung cancer 36

1.4.1 Landmark papers and studies in the characterisation of lung cancer volatiles 36

1.4.1.1 In-vitro studies 36
1.4.1.2 In-vivo studies 39
1.5 Fieldable devices – an alternative to imaging and biopsy 41
1.5.1 Mass spectrometry as a fieldable device 44
1.5.2 Ion mobility as a fieldable device 44
1.6 Data processing 46
1.6.1 Current methods in data processing 47
1.6.1.1 Mass spectrometry data processing 47
1.6.1.1.1 Deconvolution of chromatographic peaks 47
1.6.1.1.2 Identification of compounds found in breath 48
1.6.1.1.3 Cataloguing of compounds found in breath 49
1.6.1.1.4 Identification of potential biomarkers of lung disease 49
1.6.1.2 Ion mobility spectrometry data processing 50
1.7 Research objectives 51
2 An overview of breath analysis, a multidisciplinary combination of ethics, advanced instrumentation, *in-vivo* sampling and sophisticated data processing 52
2.1 Introduction 52
2.2 Experimental design 53
2.3 Breath sampling 54
2.3.1 A brief history 54
2.3.2 Considerations in breath sampling 56
2.3.3 Sampling techniques 57
2.3.4 Trapping volatiles 59
2.3.5 The adaptive breath sampler 63
2.3.5.1 Setting up the adaptive breath sampler 69
2.4 Instrument design 70

2.4.1 Thermal desorption 71

2.4.2 Gas chromatography 73

2.4.3 Mass spectrometry 76

2.4.3.1 Principles of mass spectrometry 77

2.4.3.2 Quadrupole ion trap mass analyser 79

2.4.3.3 Considerations when using ion trap mass spectrometry for breath analysis 81

2.4.4 Differential mobility spectrometry (DMS) 82

2.4.4.1 DMS as a detector for lung disease 91

2.4.5 Ion chemistry in differential mobility spectrometry ^{63}Ni source for identified biomarkers of lung cancer 92

2.5 Clinical studies 112

2.5.1 Ethical approval 112

2.5.2 Study design 117

2.5.3 Study implementation 118

2.5.4 Sample collection and transportation 119

2.6 Data processing and analysis methods 119

2.6.1 Baseline correction 120

2.6.2 Normalisation 122

2.6.3 Deconvolution 122

2.7 Data analysis methods 122

2.7.1 Multivariate analysis 122

2.7.2 Principal Component Analysis (PCA) 122

2.7.3 Discriminant Analysis (DA) 124

3 A dual channel thermal desorption gas chromatograph with differential mobility spectrometric and mass spectrometric detection. 125

3.1 Introduction 125

3.2 Instrumentation 126

3.2.1 Overview of instrument set-up 126

3.2.2 Adsorbent trap conditioning 127

3.2.3 Outline of the instrumentation system. 127

3.2.4 Differential mobility spectrometer mount 128

3.2.5 Splitting the sample 133

3.2.5.1 Mitigation of differences in transfer line temperatures. 134

3.2.5.2 Balancing flows. 134

3.2.5.3 Calibrating the flow through the columns 135

3.2.6 System operating parameters 138

4 Processing GC-DMS data surfaces. A new Euclidian-based approach 141

4.1 Introduction 141

4.2 DMS data processing 141

4.3 Developing the approach 143

4.3.1 Euclidian distance method 143

4.3.2 Data scaling 144

4.3.3 “Equivalent value” Euclidian vectors 145

4.3.4 Five peak demonstration test set 145

4.3.4.1 Evaluation of the five peak test set 151

4.3.5 Dispersion field sweep data set 151

4.3.5.1 Comparison of the effect of using the Euclidian-based approach against a summation method with the summed method to analyse dispersion field programmed data 153

4.3.5.2 PCA of dispersion field sweep data sets 157

4.3.5.3 Dispersion field sweep data sets 160

4.4 Evaluation and testing the approach with modelled data. 160

4.4.1.1 Test 1 High intensity differences in half of the discriminant data. 163

4.4.1.2 Test 2 Medium intensity differences in a small proportion of the discriminant data. 166

4.4.1.3 Test 3 Low intensity differences in a small proportion of the discriminant data. 168

4.4.1.4 Test 4 Intensity difference in a single peak. 172

4.4.1.5 Test 5 Differences in resolution 174

4.4.1.6 Test 6 Drifting phenomenon. 176

4.4.2 Evaluation and testing with a spiked system. 178

4.5 Discussion 185

4.5.1 Comparison of the two methods when applied to a real data set 185

4.5.2 Surrogate biogenic VOC profiles 187

4.6 Conclusions 189

5 Design and implementation of a collaborative clinical campaign within the National Health Service (NHS) 191

5.1 Introduction 191

5.2 Initial discussion with the Clinical lead 192

5.3 Required documentation 192

5.3.1 Ethics documentation 193

5.3.2 Researcher documentation 193

5.4 Identification of sampling site 193

5.5 Transport logistics 193

5.6 Installation and testing of breath sampling equipment 194

5.7 Training of hospital staff 194

5.8 Changes to the initial ethically approved study plan 194

5.9 Flow diagram and Gantt chart 195

5.10 A guide to setting up a collaborative clinical study within the NHS 195

6 Exhaled Breath of Participants attending a chest clinic: a Pilot Study - An initial inspection and development of a workflow. 198

6.1 Introduction 198

6.2 Methods 200

6.2.1 Sample and information collection 200

6.2.2 Sample transport and storage 201

6.2.3 Sample selection 201

6.2.4 Sample analysis 203

6.2.5 Data processing 204

6.2.5.1 Visualisation 204

6.2.5.2 Peak examination and qualitative separation of chromatographically unresolved analytes. 208

6.2.5.3 Identification and detected volatile table generation 209

6.3 Preliminary results from participants with lung cancer 212

6.3.1 Preliminary data-processing 213

6.3.2 Data processing 213

6.3.2.1 Differing intensities in sample pairs 214

6.3.2.2 Complex unresolved peaks 225

6.3.2.3 Foreline pump contamination and siloxane artefacts and plasticisers 227

6.3.2.4 Retention time shifting 234

6.3.2.5 Addressing the complexities and identifying compounds found in the breath of volunteers diagnosed with lung cancer. 236

6.4 Discussion 237

6.4.1 Compounds found in the breath of volunteers with lung cancer 237

6.4.2 Normalisation of data sets 239

6.4.3 Manual versus automated deconvolution methods 240

6.4.4 Breath compound database 241
6.5 Conclusions 242
7 Overall discussion, conclusions and future work 244
7.1 Overall discussion 244
7.2 Conclusions 247
7.3 Future work 248
Appendix I 249
Appendix II 255
Appendix III 261
Appendix IV 271
Appendix V 275
Appendix VI 278
Appendix VII 291
References 314

“Not their application, certainly, but their principles you may; to learn is not to know; there are the learners and the learned. Memory makes the one, philosophy the other.”

L’Abbé Faria

The Count of Monte Cristo

Alexandre Dumas, 1844

1 An introduction to lung cancer, current detection methods and applicable detection technologies for fieldable devices

This thesis is based on the premise that there is a detectable difference between the breath of people with, and without, lung cancer. To study this difference a large range of techniques and skills are required. From collecting the sample through to processing the final data output from the instrumentation, it is important that each step is understood.

This introductory discussion considers the underlying concepts of this research and necessarily, includes;

- Lung cancer
- Current detection techniques
- The biochemical basis of cancer targets
- Volatile organic compounds and lung cancer
- Detection technologies
- Fieldable devices – an alternate to imaging and biopsy
- Data processing methods in VOC analysis

The fundamental techniques that were used in this research, breath sampling, thermal desorption, gas chromatography, mass spectrometry and differential mobility spectrometry will be fully described in 2, although where appropriate they will be discussed here.

1.1 Lung cancer

In order to understand lung cancer and why there is a need for its' early detection, it is important to have a basic understanding of the disease state. Cancer is a disease in which cells exhibit uncontrolled growth, invade neighbouring tissue and sometimes show

metastasis. Some, but not all cancers will form tumours. It is the invasive nature of the disease and its' tendency to metastasise that separates cancer from benign tumours and makes it so dangerous. What is important to understand is that different cancers progress in different ways and this can affect both detection and treatment. The main types of lung cancer will be discussed, with an emphasis on identification, classification and method of treatment.

1.1.1 Types of lung cancer

The term lung cancer covers several different diseases that occur in the human lung. These are classified into three groups, with mesothelioma classified in isolation, Table 1.1.

Table 1-1 Types of cancer and their group

Type of Cancer	Group
Adenocarcinoma	Non-small cell carcinoma
Squamous cell carcinoma	Non-small cell carcinoma
Large-cell carcinoma	Non-small cell carcinoma
Small cell Carcinoma	Small cell carcinoma
Mesothelioma	No applicable group

All of these different cancers can be identified, but in terms of treatment the important information is the group to which the cancer belongs. The reason mesothelioma is not placed in one of the two main groups is because it is a very different disease to any of the others that fall under the term lung cancer.

1.1.1.1 Adenocarcinoma

Adenocarcinoma can be differentiated from other cancers as it originates from glandular tissue and so can sometimes be recognised by its' attempts to form new glands.

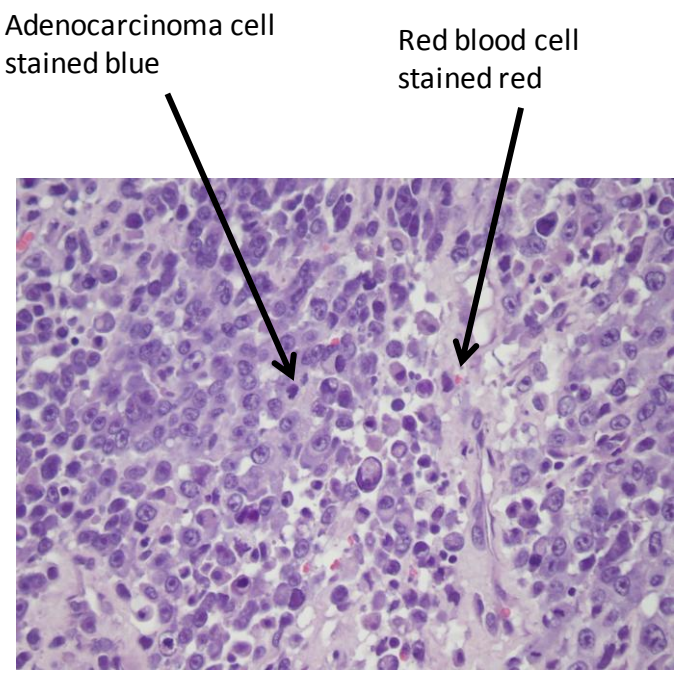


Figure 1-1 Histopathology slide of the adenocarcinoma of the lung; HE stain [1].

Figure 1-1 shows a hematoxylin and eosin stain (HE stain) of a histopathology slide of adenocarcinoma. HE staining is a useful technique in histology. It uses hemalum, a mixture of aluminium and hematoxylin, to stain the nuclei of cells dark blue. This is followed by counterstaining with a solution of eosin Y, a fluorescent red dye, which colours eosinophilic structures different shades of red. Eosinophilic structures can be composed of protein, a large part of the cytoplasm is eosinophilic and red blood cells are stained intensely red.

Adenocarcinomas can spread to the lymph nodes and eventually to the blood, which can lead to metastatic disease. Adenocarcinoma is the most common type of lung cancer seen in patients with no history of smoking and is the most common form of cancer seen in smokers. Adenocarcinoma tends to start in the periphery of the lung [2] and so occurrences of this type may be good candidates for surgery. It is a form of non-small cell lung carcinoma and so is less susceptible to radiation therapy compared to small cell lung carcinoma meaning surgery is the most likely treatment route.

1.1.1.2 Squamous cell carcinoma

Squamous cell carcinoma is closely related to a history of smoking [3]. It is a non-small cell lung carcinoma making it less responsive to radiation therapy. This means that treatment options for squamous cell carcinoma are usually surgical.

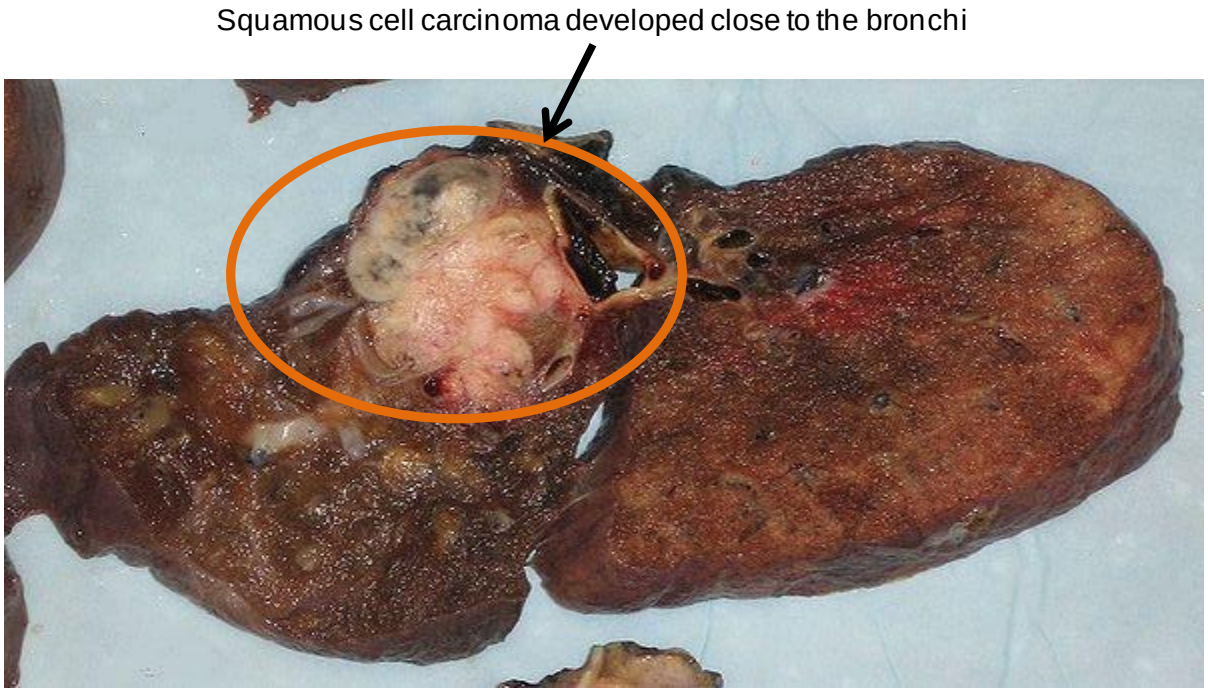


Figure 1-2 Squamous cell lung carcinoma. Adapted from [4].

This type of cancer most commonly develops in the major airways and from there can spread to the smaller airways and on to the lymph nodes and the blood, leading to metastatic disease. It may be preceded by squamous cell metaplasia, an irreversible changing of one cell type to another, which results in dysplasia, a change to an abnormal cell type, and finally forming a malignant neoplasia. Metaplasia is undetectable by x-ray. The hypothesis is that if different volatiles are released by squamous cell lung cancer cells they may be detectable by breath analysis. Earlier detection of this disease may improve survival rates and breath analysis could be a viable option as a diagnostic tool. While biopsy and analysis of the tissue allows for detection of this disease, it is a costly and invasive technique. Furthermore biopsy is usually only performed if another technique, such as CXR or CT scanning suggests the disease is present. By this point the disease has usually spread and can be fatal.

1.1.1.3 Large-cell carcinoma

Identified by its' large cells, this cancer does not form any recognisable structures, unlike adenocarcinoma. They are most commonly found in smokers and can appear either within the lung or at the periphery. Like adenocarcinoma and squamous cell carcinoma, they spread to the lymph nodes and the bloodstream and can cause metastatic disease.

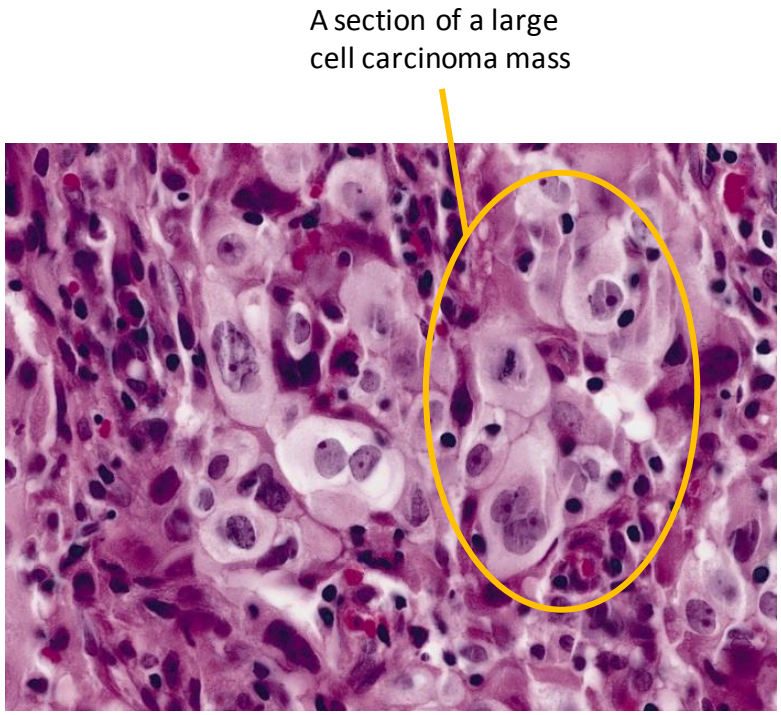


Figure 1-3 Large cell carcinoma of the lung, HE stain [4].

A diagnosis of large cell carcinoma is made when adenocarcinoma, squamous cell carcinoma and small cell carcinoma have been excluded. There are several different types of large cell lung cancer, classification of which is based on the new World Health Organisation classification of lung tumours [5]. The cells show an absence of "salt and pepper" chromatin, granular chromatin in the cell nuclei, showing up as a grainy white and blue in a HE stain. Large cell carcinomas are composed of anaplastic cells, cells which show poor differentiation between each other, exemplified by the mass highlighted in Figure 1-3.

1.1.1.4 Small-cell carcinoma (oat cell carcinoma)

Small-cell carcinoma, sometimes called oat cell carcinoma due to the cell shape, is named due to its' small, fragile cells.

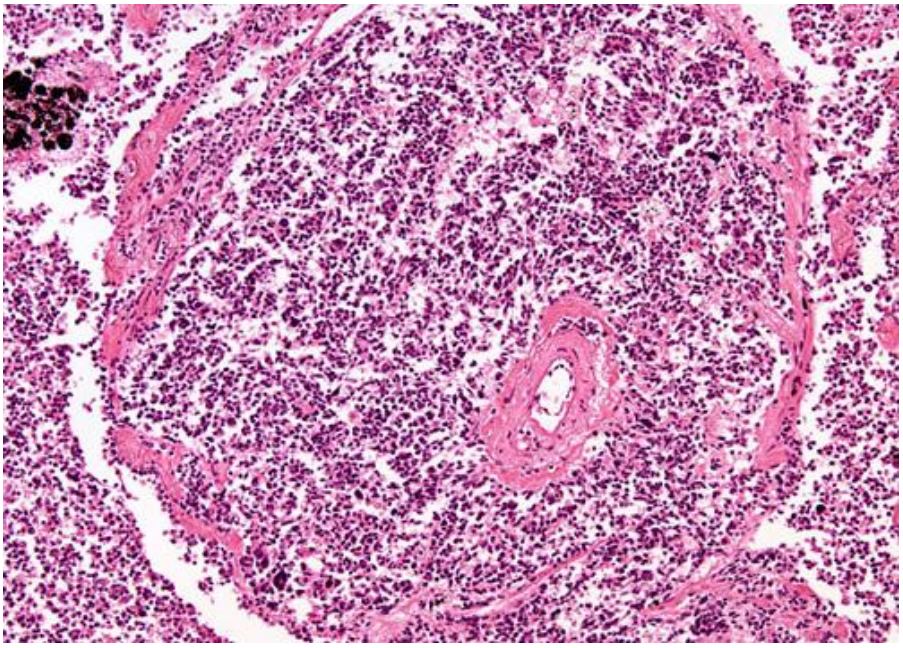


Figure 1-4 Small cell carcinoma, HE stain [6]. The cells of the small cell carcinoma are stained dark blue.

This form of lung cancer is most commonly associated with smokers and, like squamous cell carcinoma, it is most often found in the central part of the lung. Due to its' tendency to be highly malignant it tends to progress quickly and, being more metastatic than non-small cell lung cancer, spread in its' early stages. Small-cell carcinomas are thought to develop from Feyrter Cells, named for Friedrich Feyrter [7], in the bronchus.

Small cell carcinoma is usually treated with combination chemotherapy and radiotherapy when the disease is limited stage, confined to the thorax, and just chemotherapy when it is extensive stage, metastasised to other parts of the body. This is the most aggressive form of lung cancer with an average survival time for sufferers of limited stage disease of 14-20 months and approximately 20% surviving longer than 5 years and an average survival time of sufferers of extensive stage disease of 8-13 months with 1-5% surviving for 5 years or longer. This shows that catching this form of the disease in its' earliest stage would be very beneficial and a method that can detect the disease before it is visible by imaging would be highly desirable.

1.1.1.5 Mesothelioma

Mesothelioma is a tumour of the pleura, the membrane that surrounds the lungs and separates them from the chest wall. It is usually caused by exposure to asbestos and frequently causes a build up of fluid in the pleural cavity.

Figure 1-5 Mesothelioma due to asbestos exposure [8].

As mesothelioma is a cancer of the lining around the lungs a breath detector may not be applicable in its' diagnosis so it has not been considered in this work [9].

1.1.2 Causes of lung cancer

There are several causes of lung cancer, but the most widely recognised and publicised is tobacco smoking. There have been many, many studies linking tobacco smoking to lung cancer, and as smoking became more popular the number of lung cancer cases began to rise, leading deaths per one hundred thousand due to lung cancer to have more than tripled for both men and women since the 1950's [10]. It has been estimated that ninety per cent of lung cancer deaths in the U.K. are due to smoking [11].

Cessation of smoking has been shown to reduce the risk of lung cancer, with Doll and Hills' work over a 20 year study of British doctors showing that the death rate for those studied fell by 38 per cent, coinciding with over 50 per cent of them giving up smoking within the 20 year study period while the death rate for men in the whole of Britain rose by 7 per cent, coinciding with an increase in smoking [12]. A clear benefit has been shown for smokers who give up smoking early, with the risk of lung cancer rapidly falling after just 5 years of no smoking and carrying on falling for over 20 years of no smoking, to a level just above non-smokers.

It has been shown that the more cigarettes a person smokes, the more likely they are to contract lung cancer, although the major factor is how long someone has smoked for [13]. This means that regular, long term smokers are much more likely to contract lung cancer, and form the highest risk group.

Passive smoking has been indicated as increasing the chances of getting lung cancer, so the spouses/partners of smokers are at an elevated risk of contracting the disease. If their partner smokes more than 20 cigarettes a day their chances of developing lung cancer are doubled compared to someone whose spouse/partner is a non-smoker. It is estimated that 11,000 deaths a year in the U.K. are caused by passive smoking.

While smoking accounts for the large proportion of lung cancer cases, there are several other causes. Air pollution, vitamin A deficiency and exposure to asbestos have all been suggested as possibly increasing the risk of contracting lung cancer, with asbestos being the main cause of mesothelioma.

Exposure to radon gas is the second biggest cause of lung cancer behind smoking. Uranium occurs naturally in the Earth's crust and as it decays it releases radon. Radon occurs as a gas under standard conditions, it is radioactive and because of its high density compared to other gases under standard conditions it can accumulate in poorly ventilated spaces such as attics and basements. When a person enters a location with an increased radon concentration they will breathe in the gas and it can damage DNA in the cells it passes through. This can lead to mutations in the cells which can then lead to cancer.

1.1.3 Prevalence of lung cancer

Lung cancer is the most common cause of death from cancer for both males and females. Twenty four per cent of UK male cancer deaths and twenty per cent of UK female cancer deaths are due to lung cancer [11]. This is coupled with extremely poor survival rates, with only twenty seven percent of male patients and thirty per cent of female patients in England and Wales surviving for one year after diagnosis and seven per cent of male patients and nine per cent of female patients surviving for more than five years [14].

Lung cancer is a disease of the developed world, with people living in the U.K. having the highest risk of developing lung cancer. People living in Africa have the lowest risk of developing lung cancer. There are approximately 34,500 deaths in the UK due to lung cancer each year [14].

Estimates for lung cancer prevalence for the world are available through the GLOBOCAN database from CANCER*Mondial* [15] which allows users to see the cancer statistics by country for all cancer types. A world map, Figure 1-6, generated from this shows the different mortality statistics using the standardised mortality rate (SMR).

The SMR is calculated according to:

$$SMR = \frac{O}{E} * 100 \qquad \textbf{Equation 1-1}$$

Where *O* is the total number of deaths observed in a specific population and *E* is the expected number of deaths if the population was subject to a standard set of age-specific

rates. In the case of the SMR E is the age specific mortality rates for the World. The SMR is expressed as a percentage of *E* and this means that the SMR can be used to compare a specific population to the World.

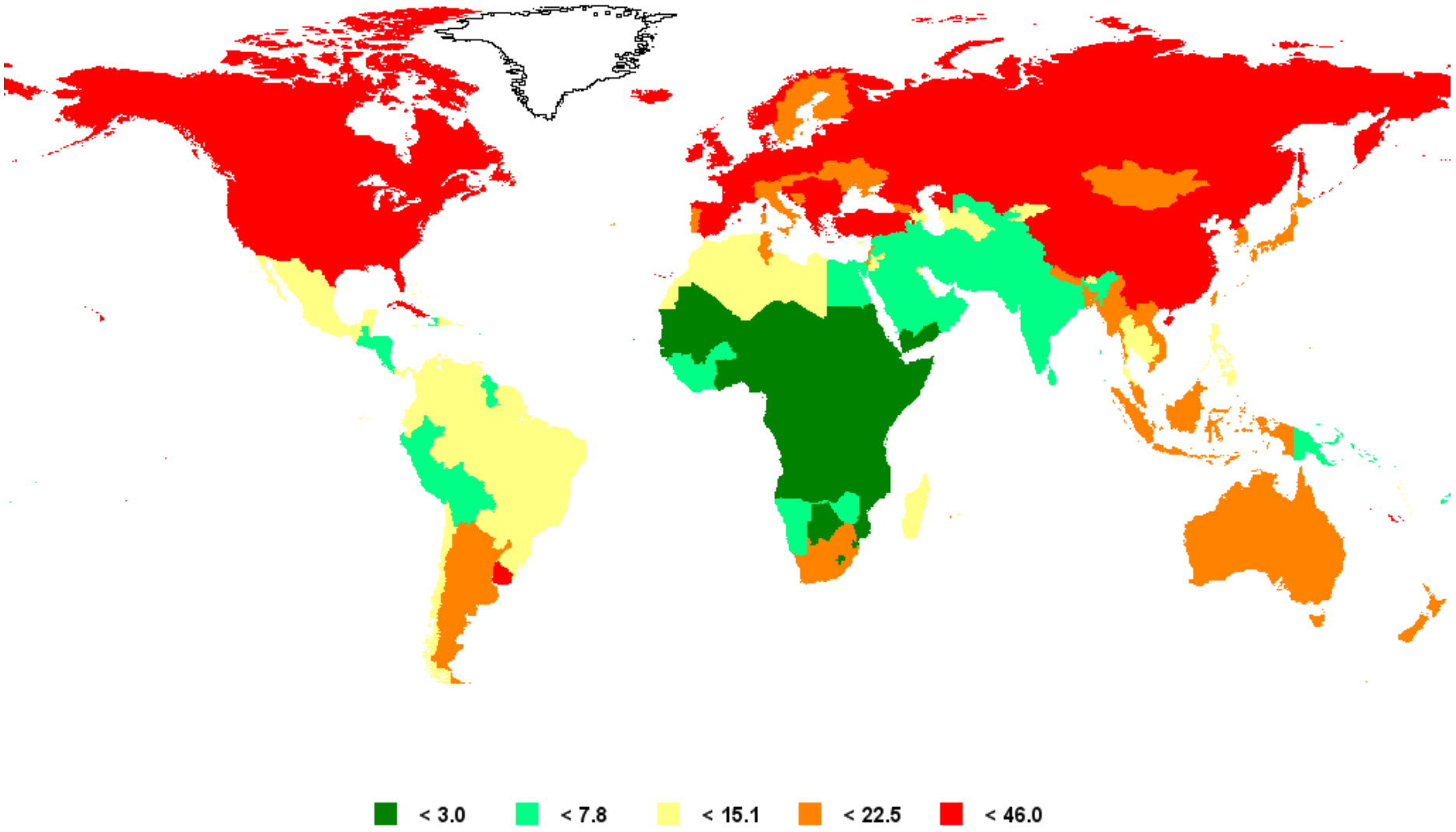


Figure 1-6 Maps coloured according to the SMR calculation to show the rate of lung cancer death in every country compared to the world average.

Figure 1-6 shows a shaded map giving the SMR for each country according to the colour key shown. From this map it is clear to see that, on the whole, the mortality rates are higher in the more developed regions of the world. The enlarged map of Europe shows the large differences in the morality rates between European countries. The statistics for Africa reveal a wide range of mortality rates that tend to follow the GNP trends seen at a Global level. But even so it should be noted that only the highest rates of mortality are above the World average.

While lung cancer is mainly a disease of the developed world, the prevalence of the disease is beginning to spread. The World Health Organization (WHO) gives the expected increase in lung cancer incidence rates for different regions of the world. These rates have been challenged as underestimating the problem as cancer registration in developing areas is poor so many are extrapolated from neighbouring countries. Work by Nawi *et. al.* Suggests that the spread has been underestimated by as much as two point six times the 2003 WHO in sub-Saharan Africa [16]. This indicates that the incidence of lung cancer will steadily increase in areas where healthcare can be poor, and low-cost fast detection methods would be of a huge benefit here.

Men are more likely to develop lung cancer than women with a rate of sixty point eight per one hundred thousand of the population compared to thirty seven point one per one hundred thousand of the population. Most lung cancer suffers are sixty or older with more than three quarters of deaths due to lung cancer occurring over the age of sixty five years old [14].

1.2 Current lung cancer diagnosis techniques

The detection of lung cancer is a multi-disciplinary and multi-parameter procedure, and there is no single diagnostic test. There are several different diagnostic tests used to determine if a patient has lung cancer. These diagnostic tests are:

- Spirometry,
- Bronchoscopy,

- Chest x-ray (CXR)
- Computed tomography (CT) scan,
- Positron emission tomography (PET) scan and,
- Biopsy.

1.2.1 Spirometry

There are many different spriometric tests. The two most common tests are: the forced vital capacity test (FVC); and, the forced expiratory volume in 1 second (FEV_1). The FVC test measures the volume of air, in litres, that can be forcibly expired after full inspiration and the FEV_1 test measures the maximum volume of air, in litres, that can be forcibly expired during the first second of expiration. Calculating the ratio of FEV_1 to FVC can give a very good indication of a persons' lung function. A value of 0.75 - 0.80 is expected for healthy adults while a value of less than 0.75 is expected for a person with obstructive respiratory diseases such as COPD or asthma; these diseases increase the resistance to higher flows in the lung . It is helpful to note that certain diseases can cause a reduction in both the FVC and the FEV_1, which would result in a normal ratio between the two. While these tests are a good indicator of whether a person has impaired lung function, they cannot diagnose specific lung disease, only that lung function is impaired. To diagnose specific diseases other diagnostic tests are required.

1.2.2 Chest X-ray (CXR) and X-ray computed tomography (CT) scan

It is possible to identify tumours within the lungs by taking a Chest X-ray (CXR). There are many different things that can be identified using a CXR, although it can be difficult to conclusively identify a tumour. The simplicity of a CXR procedure means that it is usually the first diagnostic test used when lung cancer is suspected. Where the facilities are available, X-ray computed tomography (CT) scanning is the usual next step if an abnormality is seen in a CXR. X-ray facilities are expensive to both install and maintain. They require skilled staff to operate and maintain them and they require a doctor to interpret them. They also require power services and a dedicated room in a building. This makes them unpractical for poor

communities in developing countries, particularly in rural areas with poor transportation links to local hospitals. A cheap, portable, point of care detector that requires minimal training and maintenance would be of real benefit in regions where x-ray facilities are too costly.

CT scanning is an x-ray technique that can generate three dimensional images by processing a large amount of two dimensional images based on the principles of tomography. It is more useful than CXR as it can give a better view of the chest and allow for a better assessment of the analysis. Again, it cannot always give a conclusive diagnosis, and in this case a more invasive technique is needed to confirm the diagnosis of lung cancer. As with x-ray facilities, CT scanners are expensive and require trained personnel, power services, expensive consumables and a dedicated room or facility. This again makes them unpractical for many places where healthcare is poor.

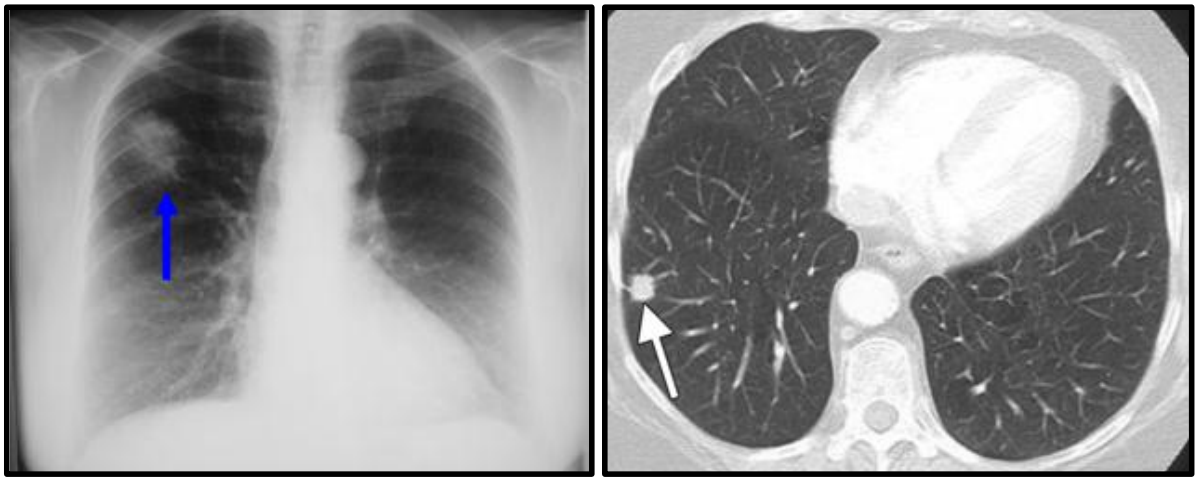


Figure 1-7 On the left is an abnormal Chest X-Ray showing a cancer in the patient's right lung (blue arrow) and on the right is CT scan shows dominant pulmonary nodule (white arrow) in right lower lobe that proved at pathology to be adenocarcinoma (adapted from [17] and [18]).

1.2.3 Bronchoscopy

Bronchoscopy is a visualisation technique used to inspect the airways. The instrument used is a bronchoscope; a device that allows the operator to see the inside of the lungs. This means the operator can examine the airways for abnormalities such as tumours, bleeding, inflammation or foreign body.

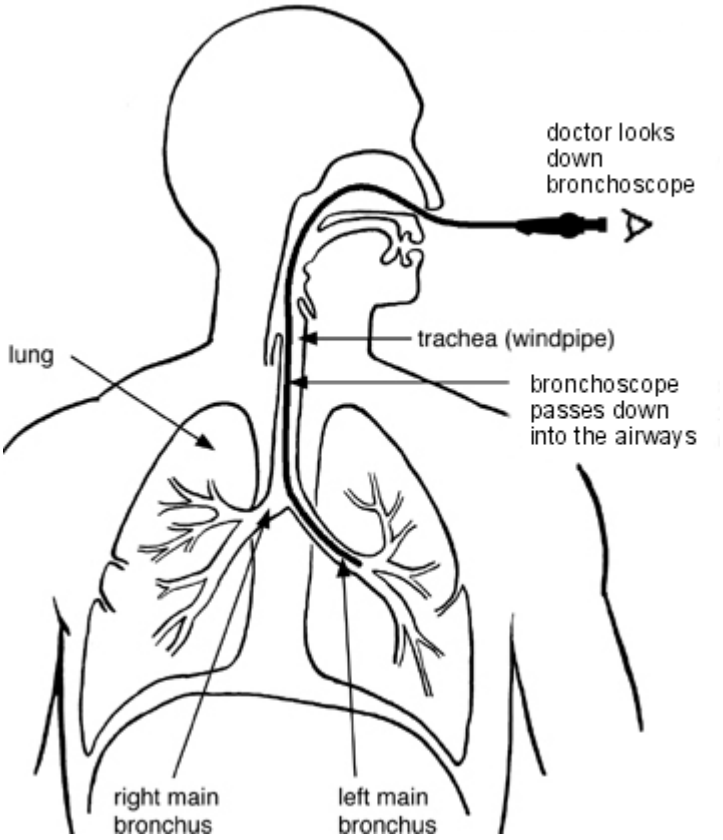


Figure 1-8 Schematic of a bronchoscopy [19].

The bronchoscope will only allow the operator to see anything within the airways, which means that, in the case of lung cancer, only tumours that have formed within the airways can be seen. A bronchoscopy can be a useful diagnostic test to confirm findings from less invasive tests, such as CT scan, PET scan or chest x-ray. A skilled practitioner is required to perform a bronchoscopy along with sterile conditions. The bronchoscope is expensive and must be cleaned thoroughly between uses. A portable detector could be fitted with inexpensive one-use mouth pieces. This would remove the need for extensive sterilisation of equipment.

1.2.4 Biopsy

A biopsy is the removal of tissue for either the eradication of an unknown, an exclusion biopsy, or the removal of a small amount of tissue from an unknown to aid diagnosis, an inclusion biopsy. In the case of lung cancer, the biopsy is the most conclusive diagnostic test that can be performed for suspected lung cancer. It is, however, an invasive test and relies

on the tumour being of a sufficient size for the biopsy to be successful. This requires a skilled practitioner, expensive equipment and sterile conditions. As it is an extremely invasive technique it is only used to confirm findings from previous imaging diagnostics. As imaging techniques cannot detect many early stage cancers and require the patient to present themselves for testing it is likely that the disease will have progressed to a later stage by the time it is detected. This reduces the chances that treatment will be successful.

1.2.5 The need for better diagnostic techniques

The current diagnostic techniques really on imaging methods that are unlikely to detect early stage lung cancer as a starting point. This means that by the time the disease is detected it is likely that it has spread and cannot be treated. This is exemplified in the poor prognosis for most lung cancer sufferers. The statistics also show that the earlier the diagnosis can be made the more likely it is that treatment will be successful and the longer the sufferer is likely to survive. This highlights the need for better detection techniques that can diagnose the disease at an earlier stage.

Diagnosis of lung cancer also relies on the sufferer presenting themselves for diagnosis. As smoking is the main cause of lung cancer, and is also a cause of many other respiratory diseases that have similar symptoms, it may be some time from the initial onset of the disease until the sufferer is aware of the onset of symptoms and presents themselves for diagnosis. This diagnosis then relies on techniques that are expensive and require skilled operators and interpreters of the data to successfully diagnose the disease. This is a limited work flow that can result in a disease being present for an extended period before diagnosis, increasing the likelihood that it will be fatal with a short time frame.

A cheap, portable, point of care detector that can be operated and maintained with the minimal amount of training with the ability to detect disease at an earlier stage would be of major benefit in the diagnosis and treatment of lung cancer. The possibility of people at a high risk of lung cancer using the device regularly to monitor their health could be realised if all these criteria are met. The aim of this work is to establish a basis for this hypothesis, showing that cheap detection methods are applicable to the detection of lung cancer.

1.3 Biochemical basis of cancer targets

To provide a biochemical basis for the biomarkers of cancer it is necessary to understand how cancer cells differ from normal cells. It is these differences that are expected to provide biomarkers of the disease that can be measured and used to develop diagnostic techniques that are capable of detecting the disease earlier than the current methods employed.

Cancer cells do not function in the same way as normal cells. Cells follow a cell cycle which is divided into two main phases, the interphase and the mitotic phase, as shown in Figure 1-9.

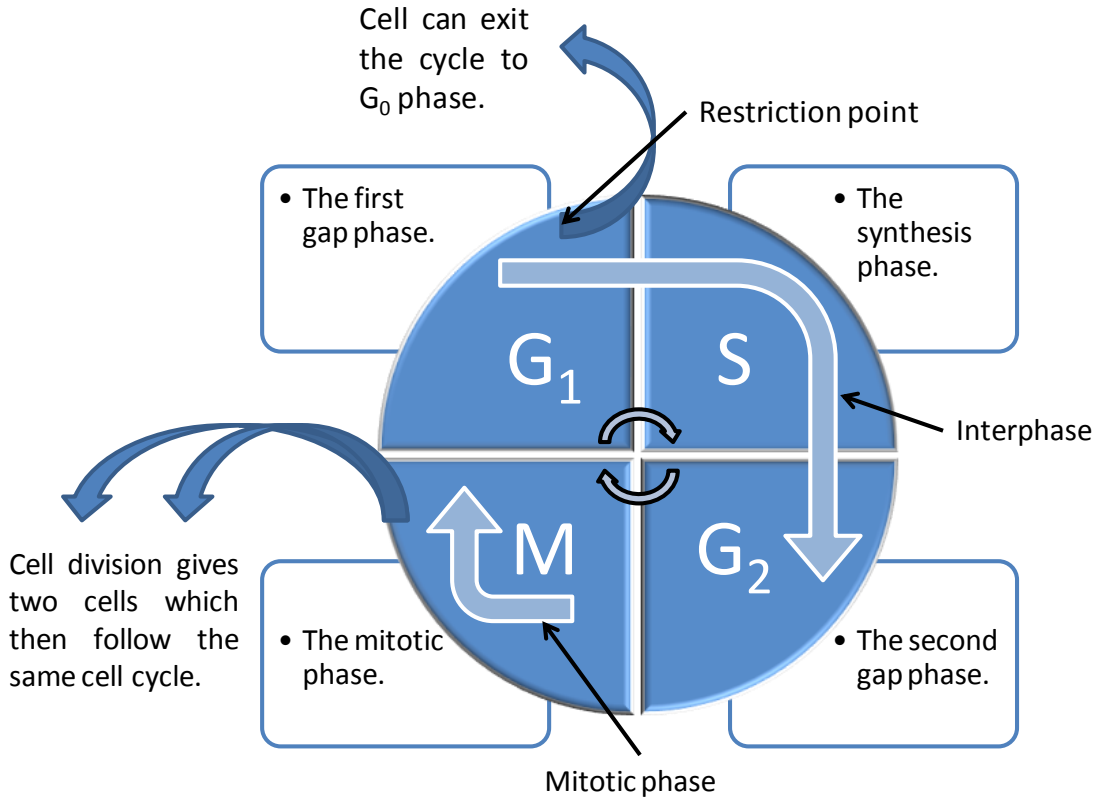


Figure 1-9 The cell cycle. Showing the four main stages: G1, the first gap phase. S, the synthesis stage. G2, the second gap phase and M, the mitotic phase.

The cell cycle, shown in Figure 1-9, is divided into two main stages;

1. The interphase. This phase can be divided into three growth sub-phases;

    i. The first gap phase, G_1. During this phase the cell is growing by synthesizing proteins to produce cytoplasmic organelles.

ii. The synthesis phase, S. During this phase, as well as growing in the same way as it is during G_1, the cell copies its chromosomes ready for cell division.

iii. The second gap phase, G_2. During this phase the cell is growing in the same way it grows in the G_1 and S sub-phases.

2. The mitotic phase. It is during this phase that mitosis divides the nucleus and cytokinesis divides the cytoplasm, creating two cells.

A population of cells increases as the cells divide. Normal cells follow a phenomenon called density–dependent inhibition of cell division. When normal cells reach a specific density they stop dividing and this can occur during the G_1 phase. Just before the cell cycle passes from G_1 to S a restriction point is reached. It is at this restriction point when it is decided if the cell will continue through the cycle and divide or if it will exit the cell cycle to the G_0 phase, the non-dividing phase. Whether a cell passes beyond the restriction point in the cell cycle is dependent on growth and nutritional status, the density of the cell population and the state of development the cell is in. The transition between phases is controlled by protein kinases which are in turn controlled by cyclins.

Cancer cells divide excessively as they do not respond to the bodies' normal cell cycle control methods. These cells can invade other tissues and, if allowed to continue, kill the organism they are present in. Cancer cells grown in culture will continue to divide and not enter the G_0 phase as long as they have sufficient nutrients. An example of this is the HeLa cell culture, a culture grown from a cervical biopsy of Henrietta Lacks taken in 1951 that is still reproducing.

Cancer cells are cells whose genes have undergone mutation, which causes them to ignore the control mechanisms that control their growth and division. The major cause of these mutations that cause lung cancer is the chemicals in tobacco smoke. Most cells that mutate are destroyed by the immune system, however if a cancerous cell is not destroyed it may proliferate and form a mass of cancer cells called a tumour. This tumour can be either

benign or malignant. Benign tumours can usually be removed by surgery and do not generally cause problems. Malignant tumours are tumours that can invade and impair the function of organs. It is the presence of malignant tumours that is classified as a cancerous disease. Cancer cells that form malignant tumours differ from normal cells in other ways as well as their uncontrolled proliferation. Differences in the cell surfaces allow them to detach and invade other tissues and separate from the tumour and enter the lymph and blood vessels. This transports the cells to other areas of the body and result in tumours growing in different tissue than the original tumour. This process is referred to as metastasis [20].

The metabolism of cancer cells may be different to that of normal cells and it is different volatile metabolites in cancer sufferers that can be used as biomarkers of disease.

1.3.1 Metabolism and metabolites

Metabolism is all an organisms' chemical processes. A cells' metabolism is made up of thousand of reactions that occur in the cell. Metabolism is specific to each cell type, although many of the same reactions will occur in different cell types. These reactions are arranged in "pathways", a series of reactions that consecutively alter a molecule. These pathways are controlled by enzymes which accept specific substrates and convert them so they can be accepted by the next enzyme in the pathway. The pathways are used to regulate the material and energy resources of the cell. The regulation of energy is fundamental to all metabolic processes and must be managed by the cell. Metabolic pathways can be sub-divided into catabolic and anabolic pathways, energy producing and energy using pathways. It is these metabolic pathways that produce the volatile organic compounds that can be measured by breath analysis.

1.3.1.1 Catabolism, catabolites and reactive oxygen species (ROS)

Catabolism is the metabolic pathways that produce energy and this sub-set of metabolic pathways can be classified as catabolic pathways. Any pathway which results in the release of energy is a catabolic pathway and they involve the breakdown of larger molecules into smaller molecules to produce energy. Cellular respiration is the major source of energy production in a cell. In this process oxygen is used as a reactant along with complex organic compounds. Fats, carbohydrates and proteins are all used in this process which produces

carbon dioxide, water and heat which is used to drive the production of adenosine tri-phosphate. Adenosine tri-phosphate is the main source of energy that drives cellular processes and this means its production and regulation is essential. The by-products of catabolism are referred to as catabolites and, as much of catabolism involves the metabolism of oxygen, many of these are reactive oxygen species (ROS). ROS are thought to play an important role in lung cancer based on the VOCs observed in breath samples from lung cancer sufferers. ROS are responsible for lipid peroxidation, the breakdown of polyunsaturated fatty acids into hydrocarbons such as pentane and ethane and aldehydes. These compounds pass from the cell into the blood and are excreted from the body via the transfer processes in the alveoli in the lungs. A possible reaction scheme showing lipid peroxidation by ROS with possible reaction products is shown in Figure 1-10.

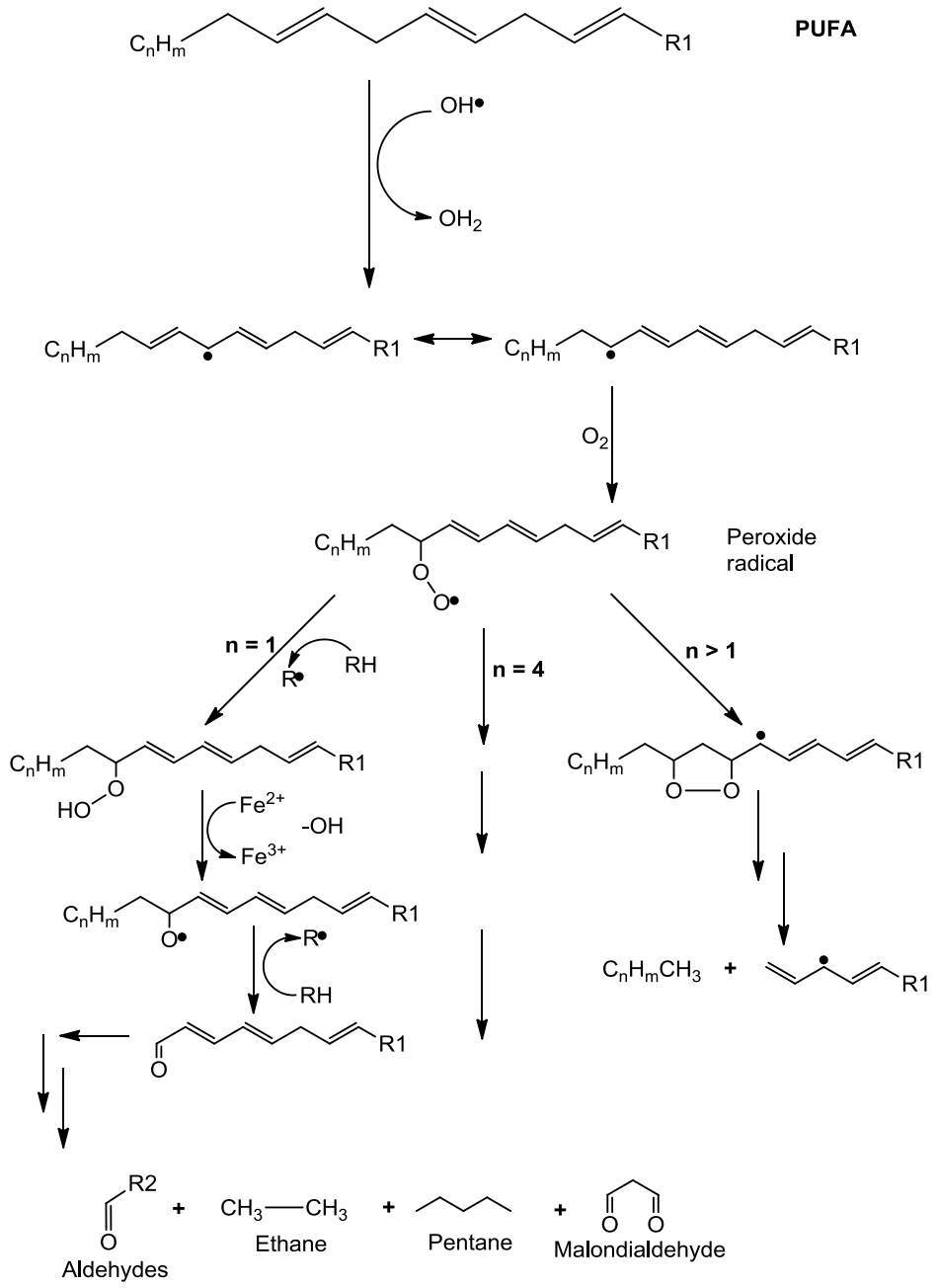


Figure 1-10 Free radical mediated lipid peroxidation: Possible reactions and reaction products. Taken from Miekisch *et. al.* [21] R = C_nH_m, R1 = COOH, R2 = C_nH_m

The role of oxygen free radicals in cancer development [22, 23] could explain an increase in ROS and this could explain an increase in the degradation products of lipid peroxidation.

1.3.1.2 Anabolism and anabolites

Anabolism is the opposite of catabolism. Anabolism uses energy to construct molecules from smaller units. Any metabolic pathway that uses energy can be classified as an anabolic

pathway. It is these pathways that are involved in the growth of cells. Different anabolic pathways exist in different cells depending on what the cell is required to do. It is these processes that will change in cancer cells and produce metabolites that are different to the metabolites found in normal cells. Identification of metabolites that are unique to or over-expressed by cancerous cells will provide biomarkers for lung cancer.

1.4 Volatile organic compounds and lung cancer

Volatile organic compounds (VOCs) are organic compounds that, under normal conditions, form vapours from their condensed state either as a liquid or a solid. This makes them strong candidates for study as disease markers. The monitoring of VOCs is a widely used technique and has been employed to detect chemical weapons, narcotics, explosives and industrial contaminants as well as monitoring air quality [24-28]. The range of organic compounds that exhibit these properties is extremely wide, so this work focuses on those found in the breath.

More than one thousand different VOCs have been observed in human breath [29], coming from both exogenous and endogenous sources. Detecting these compounds in the breath, using a range of different techniques, has been shown to allow discrimination between diseased and healthy participants in several studies [30-36]. It is the fact that volatile organic compounds can be detected in the breath that the technique of breath sampling is based on. This theme has been highlighted by several landmark papers that have progressed the research to the next stage.

1.4.1 Landmark papers and studies in the characterisation of lung cancer volatiles

1.4.1.1 In-vitro studies

Several in-vitro studies have been carried out to characterise the volatiles released by cancer cells. Chen *et. al.* [37] studied the relationship between the volatiles found in breath and cells in vitro derived from tissue removed during surgery. The breath samples were collected using a Tedlar bag system where the participant was required to take a deep breath, hold it for one second and then exhale into the bag. The authors make the statement that as there are no plants or flowers in the room the VOCs in the samples can be

considered as originating from the participant themselves. As there are many sources of VOCs from perfumes and cosmetics to cleaning products, to assume that only flowers and plants will contribute to background VOCs is a flawed approach. The cell lines were sampled using solid-phase micro-extraction (SPME), a technique in which an adsorbent is exposed to a sealed headspace above a sample and the volatiles adsorb to the adsorbent. The breath samples were extracted from the Tedlar bags using a two bag system and SPME to pre-concentrate the samples. This does not take into account any compounds which are adsorbed to the walls of the Tedlar bag which may not be removed and therefore not detected. The use of an adsorbent trap as both the sampling and pre-concentration step which eliminates the chance of these loses. The detector used in this work as a flame-ionisation detector (FID). To identify components in the breath and cell line samples a set of standards were analysed and their retention time was used to identify the same compounds in the samples. This approach limits the research to identifying compounds which have been run as standards and in this case only those previously suggested as breath biomarkers. The authors note that breath samples contain approximately 200 compounds but their research is limited to only the eleven compounds selected. While they can still show the different compounds are changing in the different samples, the detector is non-specific and it is impossible to tell if any suspected increases are due to a single or multiple compounds. They have shown that it is possible to discriminate between lung cancer sufferers, healthy controls and participants with chronic bronchitis based on a principle component analysis (PCA) plot. The small number of samples means that these results cannot be extrapolated to a larger population but the results are encouraging. However, the limitations in the sampling methodology mean that a detailed study of the meta-data is required to validate the results. Something as simple as two different operators taking the samples from a single group while wearing a different perfume. This perfume could contaminate the samples and provide separation in the PCA plot if it were mistaken for or interfered with the peak produced by one of the chosen biomarkers. By using FID as the detector instead of mass spectrometry it is impossible to know if this has happened.

Sponring *et. al.* [38] described work investigating the VOCs released by the lung cancer cell line NCI-H2087. 100 x 10^6 cells were cultured *in-vitro* in a sealed fermenter under controlled

ventilation using clean air. After sixteen hours a 200mL aliquot of the headspace was diluted 1:5 with purified dry air to combat the high humidity of the fermenter. The VOCs in this headspace were sampled using an adsorbent trap packed with Tenax TA, Carboxen 569 and Carboxen 1000. It seems like an extra complicating step to dilute with dry air when the VOCs will be extracted from the headspace as while the concentration of water may be lower as the volume of gas is increased the amount of water that is passed through the adsorbent will still be the same. The authors have noted the problems with relying on a library searching programme for identifying compounds, particularly at low levels when background interference is high and have also used the retention time of standards to confirm identification. This is a move towards making the identification more rigorous but it also relies on the compounds being run as standards for comparison. There is no mention of any deconvolution of convoluted peaks, the splitting up of peaks that are composed of one or more compounds, although it is likely that these were present in the sample analysis. The results showed that twenty two distinct compounds were detected in the headspace of the cancer cell lines which were verified by comparison with the samples. These twenty two compounds provide an initial starting place for biomarker identification. Analysis of the data showed a reduction in five compounds when compared to the medium controls. This work highlights an important point. Biomarkers may not necessarily be compounds unique to lung cancer sufferers; they may be a change in the concentration of compounds present in all samples.

Barash *et. al.* [39] sought to show that a more simple technology, gold nanoparticles, could be used to discriminate between Non-small cell lung cancer (NSCLC) versus medium. They identified 350-400 different compounds between the samples although after careful cross-comparison forty common VOCs were identified that appeared in over eighty five percent of the cancer cell lines and the control medium. The authors make no mention of how they identified the compounds in each sample but it is likely they used an automated method. If this is so then this could highlight a problem with automated methods as pointed out by Sponring *et. al.* [38], that they can incorrectly assign compounds due to interference from background ions at low concentrations. They identified fifteen VOCs that were unique to NSCLC samples compared to the medium controls. Based on these results they developed

an array of 18 chemiresistors to detect NSCLC based on its headspace. They showed that they were able to separate out four groups, NSCLC headspace, control medium headspace, lung cancer breath and healthy breath in a principle component analysis plot. This shows that a fieldable device has the potential to detect lung cancer. These results add weight to the premise behind this research. It should be noted that the breath samples were taken from participants with stage 3 and stage 4 cancers. At this stage the cancer is likely to be advanced and a detection technology would need to be able to able to detect the disease at a much earlier stage to be of benefit in diagnosis.

1.4.1.2 In-vivo studies

Breath analysis has been discussed as a diagnostic technique since it was shown to be a viable technique for measuring compounds in the breath, first by Haldane but then furthered by L. Pauling. It was Pauling who first estimated that a breath sample can contains approximately two hundred and fifty compounds [40]. The limitation of this work was that there was no pre-concentration step so compounds that appeared at low levels such as parts per million or parts per billion were not readily detectable. It was Zlatkis *et. al.* [41] who first described the use of an adsorbent to pre-concentrate breath samples. They used the adsorbent to concentrate the volatiles from fifty breaths and then analysed these. They showed that based on the profile it was possible to distinguish the difference between breath samples from people with different substances in their mouths compared to those without the substance in their mouth. The compounds that accounted for the difference were never identified.

The first study to identify compounds specifically related to lung cancer was carried out by Gordon *et. al.* [42]. This study aimed to identify differences between the VOCs found in the breath of lung cancer sufferers and healthy volunteers. They identified twenty different compounds that they used to develop a linear discriminate function that could distinguish between the breath of lung cancer sufferers and healthy controls for all twenty nine volunteers. This was the first study to show that breath analysis was a viable technique for disease detection. This work was extended in nineteen eighty eight when O'Neill *et. al.* [43] developed a computerised classification method for screening for the presence of breath

biomarkers in lung cancer. This showed that it was possible to automate the detection of lung cancer based on the volatiles seen in breath samples.

Recently several papers from Michael Phillips have built on this initial work to show that detection of lung cancer by breath analysis is a valid option as an alternative diagnostic technique, not just for lung cancer but also other respiratory diseases [33, 44-46]. This has grown the area of breath analysis, exemplified by the number of reviews recently published on the subject such as those by Miekisch and Schubert [21, 47], Chan *et. al.* [31] and Cheng and Lee [48] and even a review article on the potential of a hydrocarbon breath test as a measure of lipid peroxidation [49] as discussed in 1.3.1.1.

There has also been work on detection technology for breath analysis. Dragonieri *et. al.* [50] described an electronic nose for discrimination of patients with COPD and non-small cell lung cancer. Di Natalie *et. al.* [32] described a non-selective gas sensor array that could discriminate between lung cancer sufferers and healthy controls. These techniques are simple but they are also non-specific, they do not detect a specified molecules but a range of molecules that have an affinity for the chosen sensor. Of interest here is the work of Baumbach and co-workers who have used ion mobility as a detection technology for breath analysis and lung cancer [30, 51-53]. They have shown that it is possible to use ion mobility, a detection technique that can be tuned to specific compounds to detect volatiles from breath and discriminate between lung cancer sufferers and healthy controls [30]. This includes the development of a commercial ion mobility breath analysis device [54].

All of this work shows that the premise that there is a detectable difference in the breath of lung cancer sufferers and healthy controls is well grounded within the scientific literature. The current lack of any consensus on specific biomarkers for lung cancer needs to be addressed in the research. The identification of biological processes that may produce certain biomarkers, such as lipid peroxidiation, have been identified but this is currently limited to the action of reactive oxygen species which is non-specific. If certain lipids were identified as unique to cancer cells then it is possible that increases in particular volatile hydrocarbons would be able to discriminate lung cancer lipid peroxidation from other peroxidation, and the current literature suggests this may be the case with several papers

identifying methylated hydrocarbons as possible biomarkers for lung cancer, although this needs more rigorous investigation to confirm this. Add to this the work that has shown that relatively simple detection technologies such as electronic noses and IMS are viable as methods for breath analysis and the possibility of a fieldable device to detect lung cancer is a logical step in the future of this research.

1.5 Fieldable devices – an alternative to imaging and biopsy

The current standard for lung cancer diagnosis is the detection of the disease by one or more of the four following diagnostic techniques:

- Chest X-ray (CXR),
- X-ray computed tomography scan (CT-scan),
- Bronchoscopy and,
- Biopsy.

All of these technologies have drawbacks. CXR and CT-Scans have limits of detection of approximately 1 cm^3 [55] and cannot distinguish cancerous tissue from non-cancerous tissue. This relies on a skilled operator which experience and can still be inconclusive. Bronchoscopy can be used to visualise the airways and therefore can only detect cancerous tissue that can be seen from the airways. It also will only be carried out if the patient is suspected of having lung cancer, usually based on a CXR or CT-scan. Biopsy is an invasive detection method and will only be used to confirm that a mass found via an imaging technique is cancerous. This is because the imaging technique is needed to guide it to the suspected cancerous mass.

The main limit for all of these techniques is that the patient must first visit present themselves for diagnosis and must then be referred for the relevant tests. As many people with lung cancer also suffer from other respiratory disease, and these diseases have very similar symptoms, it can be very difficult for someone to recognise that they may be

suffering from lung cancer. The high mortality rates observed for lung cancer patients show that most people are diagnosed when the cancer is relatively advanced [56] and this can be explained by the disease only being noticed at a late stage.

The best method for lung cancer detection is regular diagnostic testing of high risk patients but this is impractical with the current diagnostic techniques. All of them are expensive, required a specialised setting and services and highly trained personnel to operate. This is a particular problem in rural areas and developing countries where the required diagnostic instrumentation is either too expensive or only located in urban areas which many people do not have access to. Add to this the limitations of these diagnostic techniques, as described in Section 1.2, and this shows that the current methods for detection of lung cancer are far from perfect.

The perfect solution would be a device that is cheap, portable, sensitive and easy to use. A device such as the fictional "tricorder" in the Star Trek series is an example of a fieldable device. While this device is fictional there are real life examples of this type of device, one of which is particularly well known, the breathaylser.

While Anstie made the first documented observation of alcohol being excreted in the breath [57], is was Emil Bogens' 1927 paper [58] that made the first comparison between the level of alcohol within the breath and the blood, stating that "...the concentration of alcohol in the breath approaches a fairly constant relationship to that in the blood, since it passes through the lungs very easily." The following work of Liljestrand and Linde [59] showed that 1 mL of blood contained the same amount of alcohol as 2 litres of exhaled air at 31 ^{o}C. Testing for alcohol intoxication is a perfect example where an unintrusive, fieldable device can replace a more intrusive test, in this case taking a blood sample. The logistical problems of taking a blood sample compared to a breath sample show that the benefits of a fieldable analysis device are huge and these are now commonplace and date back to the first device intended for use by police officers, the drunkometer [60]. While this device was the first alcohol breath tester, its' results were debatable and it was difficult to use. To combat these problems, R. F. Borkenstein developed the Breathalyser. This device was first described in the literature in 1961 [61]. The breathalyser works by passing the breath sample through a

glass ampoule containing 0.025% potassium dichromate in 50% sulfuric acid. This Concentration was calculated by Dubowski, as the initial concentration used by Borkenstein was too high for the low levels of alcohol present in breath [61]. As the breath passes through the dichromate solution, the alcohol is transformed to acetic acid. To measure the change in colour of the dichromate, the breathalyser used an incandescent light bulb arrangement on a moveable carriage that moved between a test and a reference ampoule, each with a separate photocell and blue filter. As the reaction causes a decrease in yellow colour of the dichromate, this causes a logarithmic increase in blue light transmittance according to the Beer-Lambert Law. By moving the light source to establish the null condition for the test cell, the movement is expressed in blood alcohol units (%w/v) in accordance with the movement of the light bulb carriage.

The current technology is based on the same principle that the concentration of alcohol in the blood can be measured by measuring the concentration of alcohol in the breath, but the devices have changed considerably. The alcolmeter device series developed by Lion Laboratories uses a fuel cell; an electrochemical cell that converts a fuel to an electric current. This to tuned to selectively detect the alcohol in the subjects breath. These devices show that fieldable devices are a viable option for breath analysis.

The detection technology used must be viable for the application. The technology must be capable of analysing VOCs in-situ. As it is unknown what the volatiles specific to lung cancer might be, although there are current suggestions, it needs to be capable of detecting a variety of compounds during the initial stages but also be capable of detecting a specific compound should this be required. It needs to be cheap to both purchase and run, easy to use and require as few services as possible. A device that is capable of operating as a stand-alone instrument without any external services, such as a power or gas supply that can be re-charged and then used in the field would be ideal. For this work a differential mobility spectrometer, a device based on ion mobility, was chosen and the reasons for this will be discussed, however the other option that could be considered is a mass spectrometer.

1.5.1 Mass spectrometry as a fieldable device

The main benefit of using mass spectrometry as a fieldable detection device is that it can identify unknowns. To use a mass spectrometer as a fieldable device there are several problems that need to be overcome. The size, cost and complexity of the instrumentation, along with the power required for operation, have previously limited mass spectrometers to laboratories. There has been much research into miniaturization of mass analysers [62-68], and this has led to miniature mass spectrometers being developed as the power and vacuum requirements are reduced. While this has led to a reduction in costs, the miniature turbo pumps are still expensive and this means that using mass spectrometers as fieldable devices is still an expensive option.

These devices suffer from a reduced resolution compared to laboratory devices and, while cheaper than bench-top and floor-standing equivalents, are still expensive when compared to other techniques. They are also complex and contain fragile components that make them difficult to use as a fieldable device. While further development will improve these devices and bring them closer to being a viable fieldable detector option, they are currently too expensive to be considered.

1.5.2 Ion mobility as a fieldable device

While the breathalyser is designed to measure a single compound, ion mobility devices can monitor a wide range of different compounds simultaneously. Some of the areas where these devices are most commonly employed are in the detection of explosives, chemical weapons, narcotics and biological agents, several of which are usually abbreviated to CBRN (chemical, biological, radiological and nuclear). Several companies exist that make detectors for a variety of different applications. Smiths Detection makes a wide variety of detectors, from desktop detectors commonly used at airport security, to handheld devices for onsite detection and walkthrough detectors for security check-points [69-74]. General Dynamics' JUNO detector is a small, personal detector that can identify a wide variety of chemical agents [75]. Environics is a "...Worldwide provider of CBRN protection solutions", that manufactures portable, and fixed CBRN systems and specific biological detection systems [76]. These devices are designed to detect a particular set of molecules from a sample. Some of these devices, such as the JUNO, are constant air samplers which continuously

sample air through the device and analyse it for the specific compounds. These types of device are common as personal detection devices as they allow people to move about possibly contaminated areas while continuously alerting them to possible dangers. Desktop devices such as the IONSCAN systems from Smiths Detection are commonly found in airports and other places where the threat of people transporting narcotics, explosives or biological agents is raised. Hand-held detection devices, such as the ChemPro100i from Environics, are a portable detector that can be used on demand in any location to detect the specified compounds.

These devices can also be used to monitor air in hazardous locations. The work by Eiceman et. al. [24] shows how an onsite IMS detection device can be used to monitor nicotine in the air. The major uptake of IMS devices has been the defence market, which has seen 10,000 devices put into use in airports worldwide and 50,000 devices currently being used by the US military [77]. While these may be the main areas of interest in IMS research, there is a growing trend to extend the scope to which these devices are used. They are finding a place in the pharmaceutical industry [78] and research has been carried out using them as detectors in breath analysis [30, 51, 52].

Their use in breath analysis has huge potential benefits. As has been shown in other fields they can be developed into small, robust, portable detectors. A commercial ion mobility device for breath analysis is already available [54] and these devices can be developed further by tuning them to detect specific compounds, simplifying their operation and making them more applicable to the task.

1.6 Data processing

The processing of the data generated in breath analysis is a complex task requiring several steps. The compounds found in the breath of the participants must be identified and then these must be carefully compared to highlight compounds that distinguish between the breath of healthy and diseases participants. This can be performed using the MS data. This data can then be used to tune the ion mobility device to detect the compounds of interest.

The data collected from the DMS can be used to distinguish if there is a difference between the two groups of participants. This can be done using multivariate analysis techniques which look at the full data set to find patterns which distinguish between different groups.

There are different stages to the data analysis, which are different for each detector, which must be carried out. These are;

- MS data
    - Deconvolution of chromatographic data
    - Identification of compounds found in the breath
    - Cataloguing of compounds according to the diagnosis of the participant
    - Identification of potential biomarkers
- DMS data
    - Conversion of data to acceptable format for multivariate analysis
    - Application of multivariate analysis techniques to determine if differences between groups are observed
    - Identification of data regions where differences are observed for device tuning

1.6.1 Current methods in data processing

This section seeks to provide a brief overview of the data processing techniques that are used in the area of breath research carried out with mass spectrometry and ion mobility analysis methods.

1.6.1.1 Mass spectrometry data processing

1.6.1.1.1 Deconvolution of chromatographic peaks

Deconvolution of peaks is when two or more analytes that co-elute from a separation medium are separated after detection. In terms of EI mass spectrometry this involves distinguishing which ions under the peak belong to which analyte. This is a difficult process and requires either a skilled operator to separate which ions belong to which analyte or reliance on a computer programme which does this automatically.

Both techniques have advantage and disadvantages. Manual deconvolution requires a skilled operator to decide which ions under a single peak belong to which analytes. This is a long and laborious task. The operator must have the ability to recognise which ions belong to which component but it does allow the operator to manually decide which ions belong to which analyte to give a better analysis. The use of an automated method is much quicker than manual deconvolution. Employing this method has its pitfalls as it relies on the program to distinguish between analytes which may not always be accurate.

There are few if any discussions of deconvolution of chromatographic data relating to breath analysis in the literature. Phillips *et. al.* [36] Make references to the use of library searching programs for identification of VOCs found in breath but no mention of how the data is deconvoluted first is made. An explanation for this would be that every peak in the data set is resolved, although this seems unlikely as the method used has an analytical run time of 33.83 minutes [29] during which they initially described an average number of VOCs detected per sample of 150-200. This would result in a peak width of approximately ten seconds per compound if all of the compounds were evenly spread across the chromatogram, and as can be seen in Figure 1-11 this is not the case.

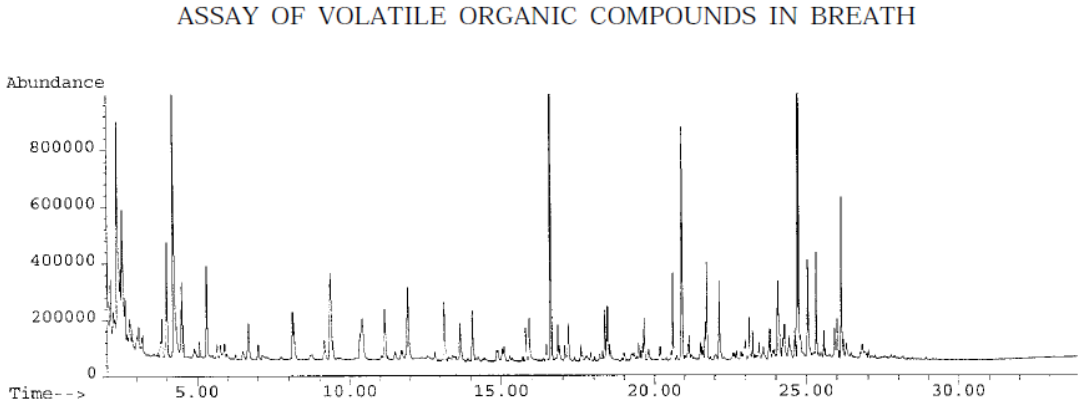


FIG. 2. Chromatogram of breath VOCs in a normal subject.

Figure 1-11 Example chromatogram produced by the analytical method described by Phillips [29]

A mixture of the two methods is the best approach but a manual method should be favoured in the initial stages of analysis. Establishing a method for this is an integral part of this work.

1.6.1.1.2 Identification of compounds found in breath

The identification of the compounds found in breath analysis is done against mass spectral libraries. As with deconvolution this can either be done manually or it can be automated. Automated methods will usually scan a mass spectral database and assign the most likely identification based on the match between your spectra and the library. This relies on the fact that the analyte is in the mass spectral library and it also relies on the fact that the top hit is actually the correct identification. As automated methods are much quicker, they are the method of choice for analyte identification. However, mass spectra are complex, particularly with the interference of background ions and co-eluting compounds and the top hit in the library may not always be correct. This means that a manual method is preferable, although it is time consuming. Being able to remove obvious contaminant ions that may have an influence over what the compound has been identified as is a large benefit. This can be a problem, especially towards the limits of detection. The main drawback of manual methods is the time it takes to assign each compound. When this needs to be done for 200 compounds per sample for 50 samples this can become very time intensive. It does have a large benefit over automated methods in that it allows you to discard an identification if you

feel that the match to the library is not correct. The compound can then be manually annotated based on the ions present, either by its compound class or by specific ions of interest.

1.6.1.1.3 Cataloguing of compounds found in breath

Careful cataloguing of the compounds is required with particular attention paid to the meta-data. Breath samples from lung cancer sufferers are difficult to obtain and this means that any samples that are obtained are extremely valuable and must be treated accordingly. Because of this it is unlikely large studies will be carried out in the early stage of this research so being able to cross-reference samples from different studies would be hugely beneficial. To make this possible careful cataloguing of the compound along with the sampling and analytical parameters and instrumentation must be carried out. Careful collection of metadata detailing aspects that may affect the volatiles in breath such as food, medication and other diseases the participant is known to be suffering from. These meta-data can then be cross-referenced against each other when determining if a particular compound is a likely biomarker of disease.

The construction of a library of volatiles found in breath, either kept in house or made available to the general research community would be very useful. Databases in other areas of research such as the HMDB [79] and METLIN [80] for metabolites and KEGG [81] for more general pathway analysis and genomic research have proved to be extremely useful tools and a similar database would be beneficial to breath analysis.

1.6.1.1.4 Identification of potential biomarkers of lung disease

The identification of potential biomarkers is the final step in the analysis of the data produced by analysing the samples using mass spectrometry. It makes use of all of the previous steps;

- Deconvolution of the data to provide the best possible data set,
- Identification of the compounds found so markers can be confirmed against standards,

- Cataloguing of the compounds to allow for cross-referencing of samples so compounds found either exclusively or at elevated or decreased levels in lung cancer sufferers can be identified.

The possibility that it is not a single compound or even compounds that are exclusive to lung cancer sufferers that will allow for differentiation of lung cancer sufferers from healthy controls must be considered. This calls for even more careful cataloguing of the samples so that any possible biomarkers can be checked against a large “back-catalogue” of samples to add weight to their identification. Once identified these compounds can be used to develop instrumentation specifically to detect them and to give some idea of the biology that is taking place to produce them.

1.6.1.2 Ion mobility spectrometry data processing

This work involves the use of a DMS and for this reason only the processing of data produced by this instrument will be discussed.

As a differential mobility spectrometer cannot provide identification of unknowns without running standards, the best method to process complex DMS data is by using a multivariate technique such as principle component analysis. The three dimensional data produced by the differential mobility spectrometer is a problem if multivariate techniques are to be used. These require two dimensional data, meaning that the data must be reduced in some way. The current methods for reducing the data are to decompose the data into its constituent wavelets [82] or to sum the data across the compensation voltage axis to reduce the data to two dimensions [46]. Both of these options reduce the information in the data set. This means that the full data set is not currently used in the data processing and this work aims to address this.

1.7 Research objectives

In priority order the research objectives were:

- To develop the dual detector mass spectrometry/differential mobility spectrometer analytical instrument set-up to increase the amount of information gained from the analysis of each sample,

- Instigate new data processing methods to allow for three dimensional data sets to be inputted into multivariate analysis tools,

- Outline a workflow for establishing a clinical trial between two research sites, an analytical facility and a National Health Service lung cancer clinic detailing the documentation, logistics and time frame required,

- Give the initial data collected from a study into participants with lung cancer, including a list of compounds found in the breath of lung cancer sufferers.

2 An overview of breath analysis, a multidisciplinary combination of ethics, advanced instrumentation, *in-vivo* sampling and sophisticated data processing

2.1 Introduction

The hypothesis of this work is that there is a detectable difference in the breath of respiratory disease sufferers when compared to healthy controls. To test this hypothesis a multidisciplinary research programme was required. Figure 2-1 introduces the different aspects of this research and how they depend through a work flow diagram.

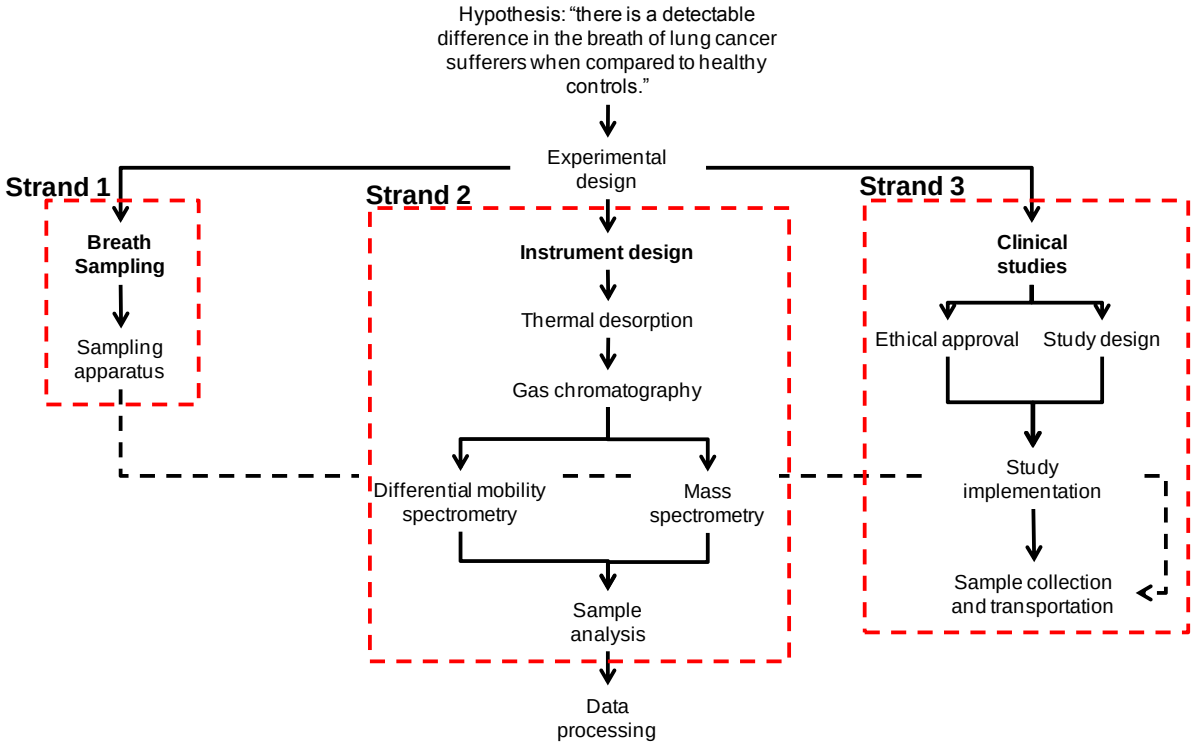


Figure 2-1 Work flow diagram showing the different aspects of the research programme. There are three main strands of the research; clinical studies, instrument design and breath sampling. These three strands are all encompassed under the experimental design. Each strand is composed of several steps listed under the main strand heading. Separate from these strands is the final step of data processing.

Each strand from Figure 2-1 will be discussed in detail, with the different aspects and how they interact with the other strands outlined. Challenges posed by the work and the solutions will be outlined along with a detailed discussion of the history and technology of the relevant steps.

2.2 Experimental design

The experimental design must take into account all aspects of the research and deliver a scientifically credible balance between the conflicts inherent in the different factors in the project's work flow. For example the sampling protocol must comply with the safeguards and restrictions necessarily specified in the project's case for ethical approval while at the same time it must be compatible with the instrumentation. Further, data analysis must form an important element of the planning stages and often designs may require one hundred plus samples with biological and machine replicates; which may not be possible due to the limitations specified in the ethics case. Finally, destructive analysis may render machine replicates impossible.

The instrument design and sample analysis must be carefully thought out. To extract the maximum amount of data possible from the samples collected a dual detector system was designed and set-up. The sample injection system used was a thermal desorber (Markes UNITY), split into two chromatography columns. One connected to a Varian quadrupole ion trap mass spectrometer and the other a Sionex S-VAC differential mobility spectrometer. This set-up allows for the DMS to be evaluated against the mass spectrometer. This means that with this set-up it is possible to determine if the DMS is capable of detecting the same changes in human breath samples as the mass spectrometer.

These different detectors generate different data outputs that require different processing approaches. The mass spectrometry data sets require deconvolution programmes to refine the chromatographic peaks and enable individual compounds in complex and poorly resolved profiles to be identified. There are no data-processing systems available in the open literature or from commercial suppliers for DMS data, so an important element of this research was the development of a new pre-processing technique for GC-DMS data.

The three strands are;

- Breath sampling
- Instrument design and,
- Clinical studies.

2.3 Breath sampling

2.3.1 A brief history

Breath sampling in its modern from has a history that stretches back for over 100 years. The first breath sampler for collecting alveolar air was described by Haldane and Priestley in 1905 [83]. This is shown in Figure 2-2.

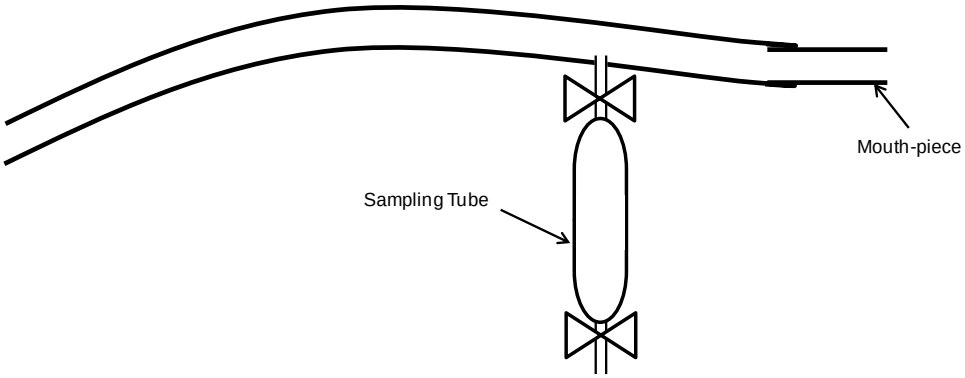


Figure 2-2 Haldane Priestley tube, described in and adapted from [83].

Initially the sampling tube is filled with mercury. The volunteer breathes normally through the mouthpiece for several minutes. At the end of a normal inspiration the volunteer exhales sharply through the tube and then blocks the mouthpiece with his tongue. To draw a breath sample the valve at the top of the tube is opened, the valve at the bottom is opened; the mercury flows out and draws the breath sample into the tube. The taps are then closed and the sample is taken for analysis. The concept of a haldane tube is still used by researchers in the field of breath sampling [84]. Breath testing for nitric oxide [85] as a marker for airway inflammation has been shown [86] and tests for hydrogen, carbon dioxide and methane as indicators in gastroenterology initially focussed breath sampling on this small range of gases. This was broadened to a wide range of volatile organic compounds

(VOCs) with the publication of work separating 200 compounds by gas chromatography in 1971 [40]. This area is now extensively researched, with a particular focus on disease biomarkers being prominent. Work identifying possible biomarkers for lung cancer has been published [31, 36] and methods for detecting lung cancer have been described [32, 87] and the idea of a "breathalyzer" to aid lung cancer detection has even been discussed in a commentary in The Lancet [88].

There are two general approaches to breath sampling: direct analysis and archival sampling, for remote analysis. Direct analysis involves the participant breathing into an analytical instrument directly, from something as simple as a breathalyser to more complex set-ups such as the Breathspec IMS instrument from G.A.S. [54]. The work by Ruzsanyi *et. al.* [52] has shown that IMS systems are a viable set up for detecting human metabolites in breath. The main problem with current direct analysis methods is that they require the participant to force expire into the instrument which can be difficult for people suffering from lung dysfunction. There are other problems with this method that can make it problematic:

- Equipment requires sterilising after use. This either means removing parts of the direct sampling system or having disposable parts that may increase costs. Infection control is essential and this means all parts that come into contact with more than one participant must be sterilised in between uses. This can be difficult with integrated instruments,
- Daily validation and maintenance of the system requires an onsite member of staff to be trained or visits from maintenance engineers,
- Installation of a dedicated sampling system requires a workspace and operators will require training,
- The initial cost of the system may be expensive, particularly if it is not heavily used and,
- Trace contamination by volatile organic compounds, and the subsequent decontamination of sensitive parts requires specialist systems (such as a vacuum oven) that may not be available on site or are expensive.

Archival sampling involves taking a sample that is analysed at a later time [35, 36, 46, 89-93]. There are several approaches to archival sampling which tend to differ in trapping method and sampling protocol. The basic premise of archival sampling is that the sample is taken at one location, stored and transported to another location. This method allows researchers who are not based in a clinical environment to collect samples at a clinical site and transport them to another facility for analysis. This combats many of the problems highlighted for direct analysis. The equipment can be taken to the lab and sterilised, the instrument can be maintained by trained operators, portable sampling equipment means that there is little to no installation required, a single system can be used at multiple sites reducing purchase costs and decontamination of sensitive components such as any face masks used can be carried out in a dedicated facility.

Sampling VOCs directly from the breath has been shown to be a useful tool for rapid detection of disease [94] and this work has led to the production of a miniaturised detection system being produced by Draper Laboratory. All this means that breath sampling could be used as a non-invasive monitoring technique in healthcare applications instead of invasive techniques such as blood sampling.

2.3.2 Considerations in breath sampling

VOCs are either endogenous or exogenous. This means that they are either a product of a process within the body or they are compounds that have entered the body, either through inhalation, absorption through the skin or consumed [95, 96], then exhaled in the breath. Exogenous compounds, studied in parallel with lung disease, can provide information into how these compounds affect people exposed to them. Endogenous VOCs can provide information on effects on metabolic pathways disease have, along with giving information of disease state and more general lung function markers. Some examples are changes in levels of Isoprene and other VOCs in COPD cases [97-99], and acetone as a marker for diabetes [100-102].

Exogenous compounds can be confounding factors in any analysis; they may influence results and cause false positives or negatives in statistical analysis.

There are other factors that should be taken into account. The concentration of VOCs is different in different places within the airways [21, 103] and so an ability to sample a particular portion of the breath profile is desirable. Patients suffering from some form of lung dysfunction may have difficulty using forced expiratory samplers and so development of sampling techniques that allow for normal breathing which can selectively sample a particular portion of the breath are desirable [104].

Human breath is a dynamic system. The Alveoli provide the location for rapid exchange between the blood and the air [105] and further VOCs can be added to this by volatile metabolism products other parts of the body such as the stomach. The dynamic range of compounds within the breath can also cause problems, with water being at a much higher concentration that any other compound found in the breath. Exogenous factors from perfumes, make-up and food can also influence the breath profile. These may also come from other people who are interacting with the subject, particularly researchers who are operating the sampling equipment. There are ways to limit these, using a full face mask and supplying purified air being effective [104], but their effects can still influence the data.

The state of any sampling equipment used must also be considered. Regular cleaning by a standard process can help eliminate contaminant build up and also standardise any contamination introduced by these processes.

These confounding factors are not always easy to control and any robust analysis must be able to take them into account to make it feasible as a point of care detection method.

2.3.3 Sampling techniques

The sampling of volatile organic compounds from breath can split into two categories in terms of the technique used;

- Forced expiry and,
- Non-forced expiry.

Forced expiry involves the volunteers forcibly expelling their breath to enable a sample to be taken. A non-forced expiry approach requires the volunteer participant to breath "normally" while samples are taken. There are four reasons why a non-forced expiry approach is preferable:

- People suffering from lung dysfunction have a poor FEV_1.

The most likely candidates for breath detection are respiratory disease suffers. People suffering from respiratory disease tend to have poor lung function, which can make forced expiration a problem. Being able to take a breath sample from a volunteer who is breathing freely is therefore desirable. It is also possible that forced expiration may exacerbate a volunteers' condition which as well as causing confounding markers to be over expressed, resulting in a poor analysis, may put participants at risk.

- Sampling of larger volumes is feasible.

For forced expiration methods it is usual to ask a volunteer to forcibly expire once to acquire a breath sample. This limits the volume sampled to the volunteers' lung capacity. Non-forced expiry methods allow the volunteer to breathe normally and selectivity sample portions of the breath profile, allowing for sampling of a theoretically infinite volume.

- Blood pH changes during forced expiration and this alters the partitioning of volatiles between the blood and the air in the lungs.

Reproducible sampling is a requirement of breath sampling and it is desirable to eliminate anything that may alter this. As it is extremely difficult to fully control forced expiration, particularly in people with respiratory dysfunction, the changes to the partitioning of the gases would not be consistent. A non-forced expiry method eliminates this problem.

- Ventilation rates change during forced expiration causing a shift in the different volatiles' equilibria between the blood and the air in the lungs.

This is another example of where forced expiration methods can change the composition of the expired gas that will be sampled. A non-forced expiration method helps to eliminate this as the volume of gas expired during normal breathing at rest is more consistent than that of forced expiration.

Forced expiration methods are usually simple to perform and the apparatus is easier to maintain, but as outlined above, the benefits of non-forced expiration methods outweigh the benefits of forced expiration methods.

2.3.4 Trapping volatiles

There are 2 methods available for trapping volatile compounds from breath samples. They can either be trapped inside a container or on an adsorbent. Some researchers have advocated using Tedlar bags which are specified for several United States Environmental Protection Agency (US EPA) methods, including methods 3, 18 and 0040. Tedlar bags are composed of Tedlar®, a polyvinyl fluoride film, Made by DuPont™. Polyvinyl fluoride is suited for sampling bags because of its' low vapour permeability, it' general resistance to stresses and strains and resistance to most chemicals. A Tedlar bag can be formed from two sheets of Tedlar and a check valve. An example of commercially available Tedlar bags can be seen in Figure 2-3.

Figure 2-3 Example of Tedlar® Bags fitted with different valves [106].

While they are suitable for air sampling they do have some associated problems [107, 108];

- Recovery of VOC from Tedlar bags requires a secondary sampling step that results in a loss of efficiency and the introduction of selectivity into the sampling process.
- They fail leading to leaks and punctures,
- Ultraviolet degradation of the sample is a possibility, although black Tedlar is available, and degradation of the Tedlar can cause contamination and,
- Adsorption of VOCs onto the walls of the bag can cause both losses from the sample and cross-contamination of the next sample if the bag is not fully cleaned before repeated use.

- Hydrolysis and oxidation artefacts can be created during sample transport and storage.

- Diffusion of water vapour and other contaminants through the walls of the bag.

Some may advocate sampling canisters which although robust suffer from many of the same drawbacks.

The other method common in breath sampling is the use of adsorbents. Adsorbent based methods involve passing the sampled breath through a bed of active adsorbent (ca 100 mg). The VOCs undergo physical adsorption onto the surface of the adsorbent and are retained. Recovery is achieved by thermal desorption and this approach has advantages over Tedlar bags;

- They are robust and easy to transport.

- They are stable and samples can be stored without significant degradation.

- Recovery of the trapped VOCs may be automated ,

- Thermal desorption provides close to 100% recovery of adsorbed VOCs from the adsorbent trap in a simple, reproducible procedure,

- Hydrophobic absorbents can e used to minimise the water content of the samples and,

- The opaque construction eliminates ultraviolet degradation.

- Volatile enrichment is possible using thermal desorption as it is a two stage process. A sampling volume of 2.5L is desorbed from the adsorbent using 200mL of helium.

This is then refocused on a cold trap and desorbed using 200µL of helium. This leads to an enhancement of approximately 10^4.

There are, however, disadvantages associated with active adsorbents:

- Breakdown of adsorbents during thermal desorption process can cause artefacts in the analysis,
- Sampled water is not easily removed prior to thermal desorption and,
- Competitive adsorption can influence which compounds are most well retained.

Of these problems the most easily controlled is competitive adsorption. As long as the break-through volume of the tube is not exceeded then competitive adsorption should not be an issue. The amount of sampled water is the major consideration, and while it may seem sensible to use a condenser arrangement in front of the trap this can remove compounds of interest at low levels and is not applicable. To combat the problem of water in the samples, careful selection of the active adsorbents is required.

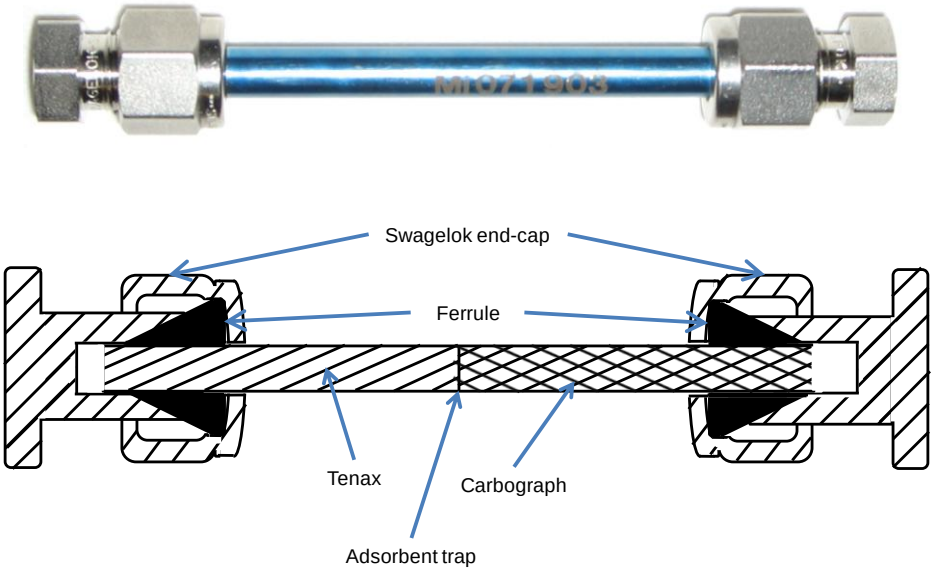


Figure 2-4 Example of an adsorbent trap fitted with end-caps supplied by Markes International and a schematic cross section of the trap showing the two different adsorbents and the end-caps fitted with PTFE ferrules.

Figure 2-4 shows an adsorbent trap of the type used for this work. These were supplied by Markes International and are inert coated stainless steel tubes packed with Tenax TA and Carbograph 1TD.

Tenax TA is a hydrophobic adsorbent with an approximate analyte range of C_7 to C_{30} with boiling points of less than one hundred degrees centigrade for apolar compounds and one hundred and fifty degrees centigrade for polar compounds. It produces a low concentration of inherent artefacts (less than one nanogram) and is inert making it suitable for labile compounds.

Carbograph 1TD is a non-specific carbon sorbent which can be used to trap a wide range of volatiles. This adsobent is hydrophobic which makes it a good choice for humid sampling, such as in breath analysis. It is generally used for trapping ketones, aldehydes, alcohols and apolar components making it a good choice for non-targeted work. It also produces a very low concentration of artefacts (less than 0.1 nanogram). Selecting the correct adsorbent for the purpose is vital as this can influence what compounds are trapped. In this work, as the compounds of interest are unknown, the adsorbents chosen were general purpose ones with hydrophobic properties.

2.3.5 The adaptive breath sampler

This work used the adaptive breath sampler that has been described previously [104]. The main features of the adaptive breath sampler are its non-forced expiry sampling method and its selective breath portion sampling ability. The adaptive breath sampler makes use of a non-vented full face mask so the volunteer is supplied with a controlled supply of medical air in an effort to control the levels of exogenous volatiles from the sample. As the air supply is constant, the volunteer is able to breathe normally while wearing the mask. Connected to this mask is an adsorbent trap via a tube assembly (shown in Figure 2-6) and a pressure sensor in the sampler control box by a silicon tube. This box contains a circuit (shown in Figure 2-7) that, along with a computer programme, monitors the volunteers breath profile and switches between two valves, both connected at one end to a pump and the other ends connected to the atmosphere for one valve and the adsorbent trap for the other. Depending on the settings for the computer programme these valves switch, drawing the breath of the

volunteer onto the adsorbent trap at specific points in the breath profile. This allows for the selective sampling of particular sections of the volunteers breath sample. A sample volume of 2.5 dm^3 was collected following previous methods developed for this sampler.

The breath sampler is composed of the following parts, which have been divided into sections:

Face-mask set-up:

- non-vented full face masks - ResMed
- PTFE T-Piece – Custom made. Schematic shown in Figure 2-5.
- One way disposable valves - Clement Clarke International Ltd Part No. 3122061 (Blue)

Adsorbent trap assembly:

- PTFE luer capillary fittings – custom made, shown in Figure 2-6.
- Steel capillary – ResTek UK
- Markes stainless steel sorbent tubes (Tenax/Carbograph 1TD) with Swagelok Brass Cap for 1/4 in. Swagelok Tube Fitting B-400-C – Markes International

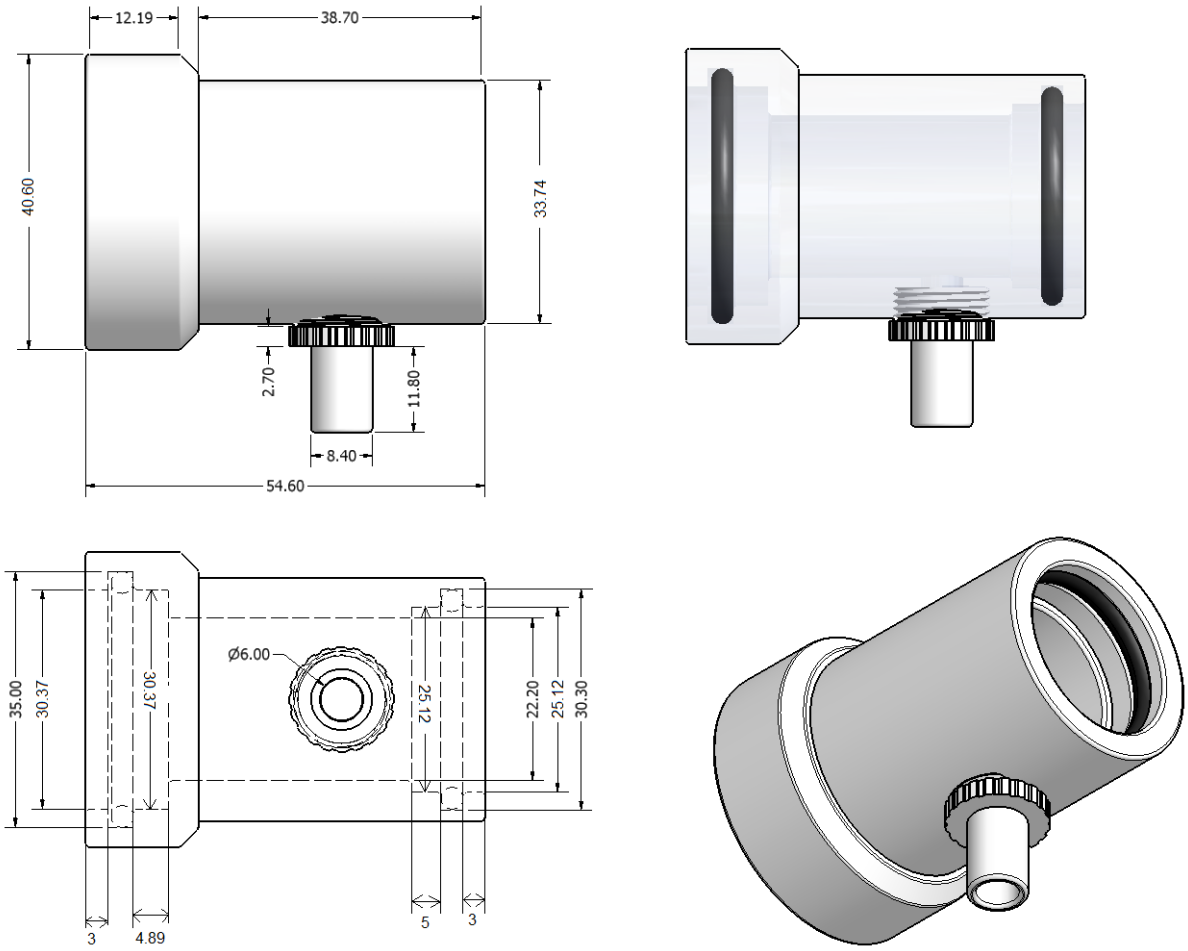


Figure 2-5 Schematic of the t-piece used to attach a one-way valve to the face mask and supply the purified medical air.

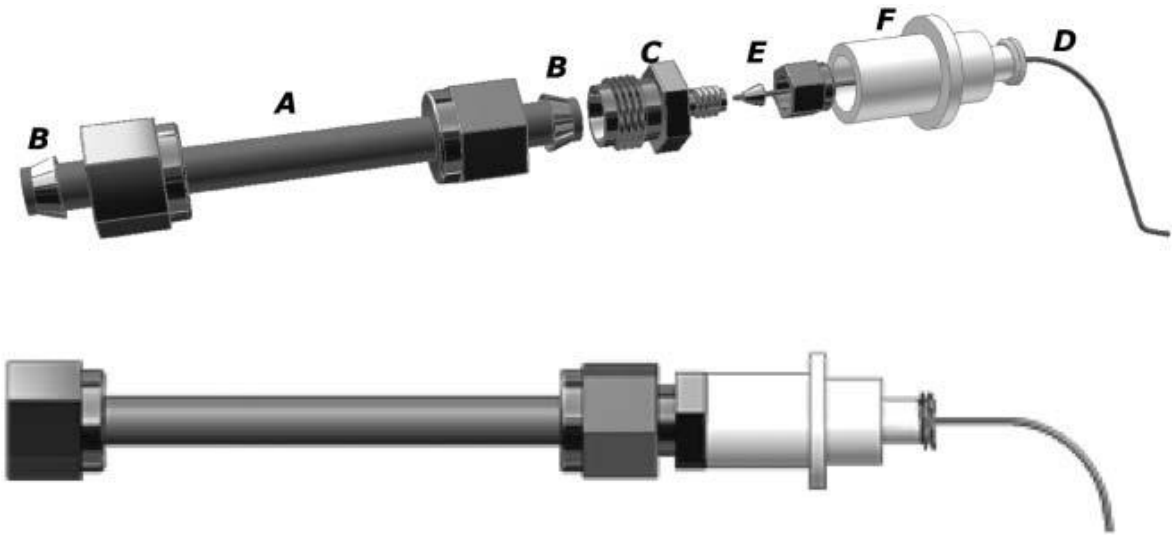


Figure 2-6 Adsorbent sampler assembly. The adsorbent trap (A) was connected to the sampling line through a 1/4'' push fit connector sealed with a PTFE ferrule, (B). The other end of the trap was fastened into a 1/4'' to 1/16'' reducing union, (C), also with a PTFE ferrule (B). A shaped 5 cm MXT passivated stainless steel column was passed through the mask connector and reducing union so it was inserted into the top of the adsorbent trap (D). A 1/16'' nut with a 0.53 mm Vespel™ ferrule sealed the capillary in place (E). The mask connector (F) was machined from PTFE with a male Luer fitting. The reducing union was held firmly in place while the capillary was threaded and positioned carefully inside the mask, before the Luer fitting secured the whole assembly. Taken from [104].

Pneumatic parts :

- 2002N-1B1-BX 2002-series (1/4") 1/4" NPT DPI auto drain BXE grade cartridge, 2002N-1B1-DX 2002-series (1/4") 1/4" NPT DPI auto drain DXE grade cartridge, 2002G-0A0 2002-series (1/4") 1/4" NPT no DPI drain plugged NO CARTRIDGE - Parker Balston Filters from KC controls
- 100-12-BX filter elements; box of 10, 100-12-DX filter elements; box of 10, CI100-12-000 2002-series (1/4") adsorbent filter element; box of 1
- Balston clamp kit C02-2091 x2
- Silicon tubing – 4.8 mm bore, 6.4 mm bore, 3.6 mm bore – RS

Gas fittings for the sampler box:

- SS Swagelok Tube Fitting, Zero Volume Reducing Union, 1/4 x 1/16 in. Tube OD Part No. SS-400-6-1ZV - Swagelok
- 2 x SS Swagelok ¼ tube x 1/8 pipe Part No. SS-100-R-4 – Swagelok
- SS Swagelok union tee 1/16 in fittings Part No. SS-100-3

Electronics:

- Sampler Box – schematic shown in Figure 2-7
- 12v DC Power plug for sampler box - Farnell
- Escort Elf personal pump - MSA
- Laptop running windows 2000 with a pcmcia card slot
- National Instruments DAQcard-6024E (for PCMCIA)
- 1.5m D25M – VHDCI M Cable

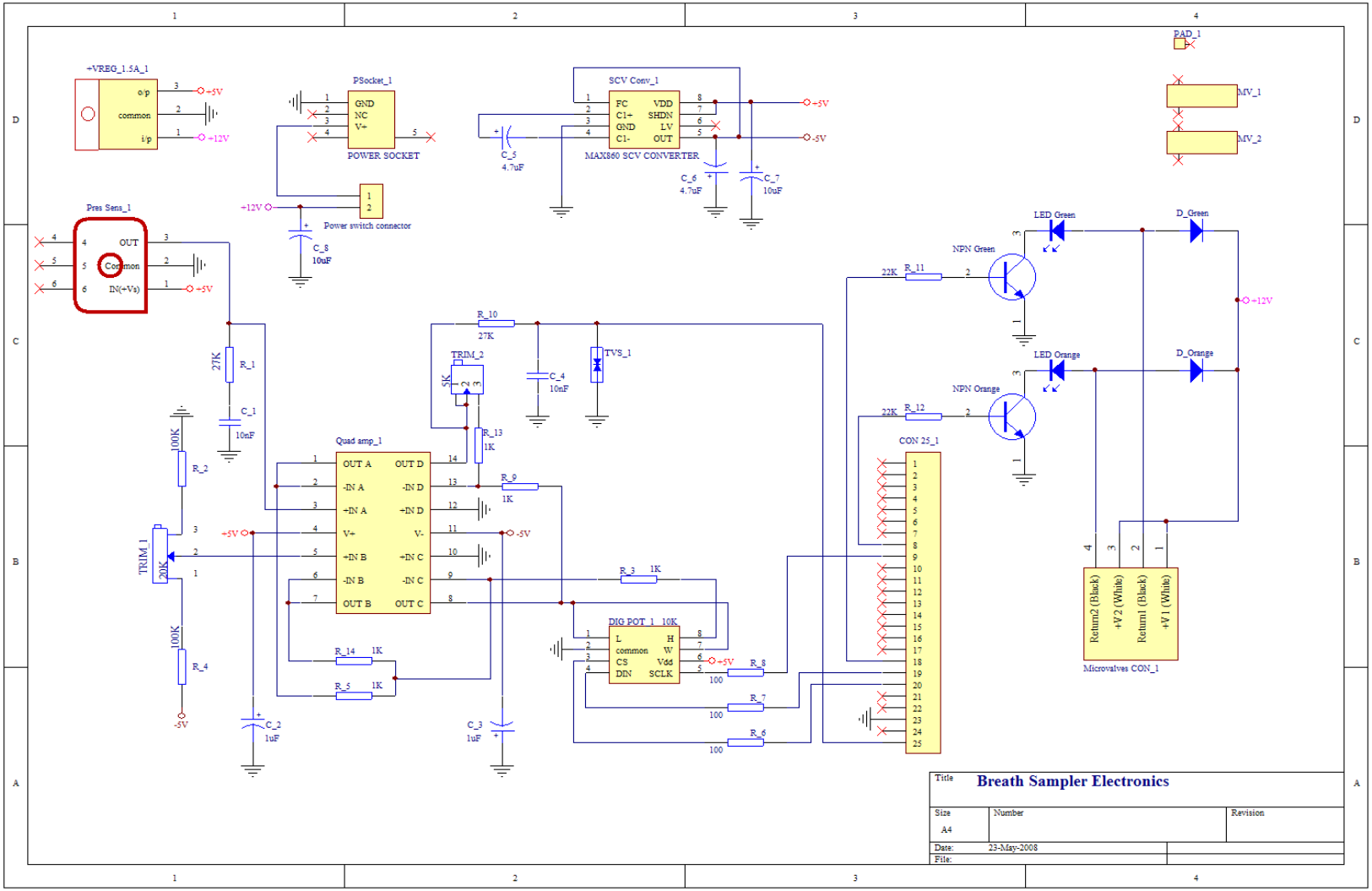


Figure 2-7 Schematic of the Breath sampler control box electronics

A schematic of the sampler set up is shown in Figure 2-8. This shows the connections between the main parts of the sampler.

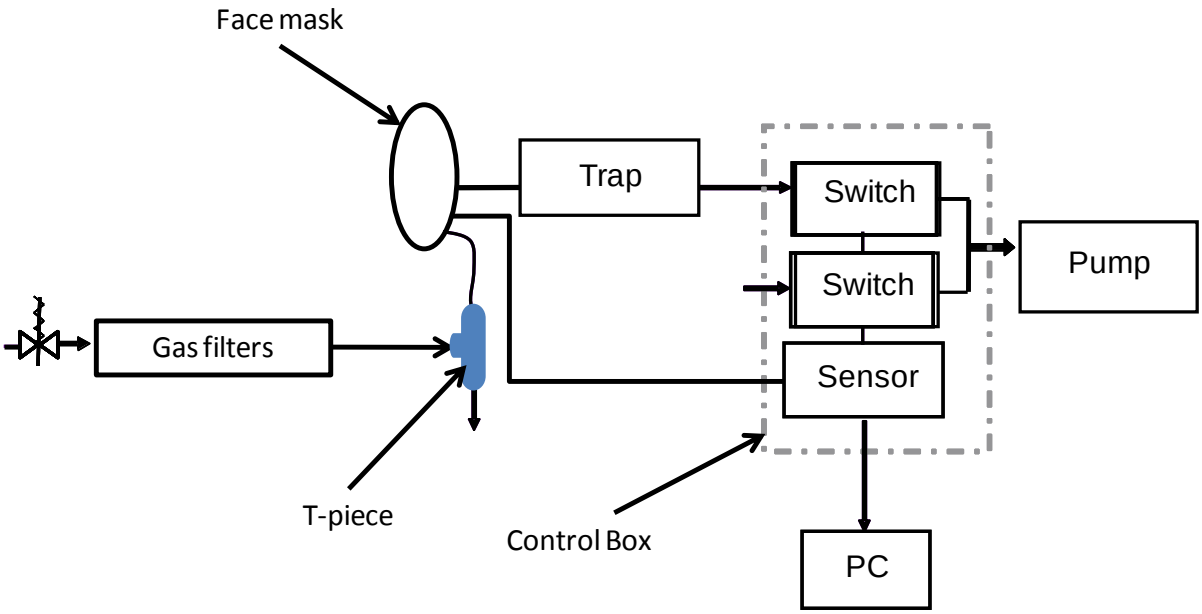


Figure 2-8 Schematic of the breath sampler set-up showing all of the major elements and their connectivity.

2.3.5.1 Setting up the adaptive breath sampler

To set the breath sampler up the following protocol should be followed:

1. The adsorbent trap assembly, shown in Figure 2-6, is connected to one of the two Luer fittings on the face mask without a tube connected.
2. A clement Clarke one-way valve is connected to the t-piece which is then connected to the face mask.
3. The Balston filters are connected to the medical air supply at the wall outlet via a 6.4 mm bore silicon tube at one end and then connected to the t-piece via another 6.4 mm bore silicon tube at the other end.
4. A 3.6 mm bore silicon tube is connected to the second Luer fitting at one end and the pressure sensor in the control box at the other.

5. The escort elf personal pump is attached to the designated pipe fitting on the control box via a 4.8 mm bore silicon tube and another 4.8 mm bore tube is connected to the other pipe fitting.

6. The sample box is powered by a 12v DC power supply and is connected to the laptop via a D25M – VHDCI M Cable to a PCMCIA card.

7. Once the mask has been fitted to the participant and they are breathing comfortably an adsorbent trap is connected to the adsorbent trap at one end and to the 4.8 mm bore silicon tube previously connected to the pipe fitting on the control box at the other end.

8. A 2.5 L sample of the participants' breath is then taken. Once this has been done the tube is removed and sealed with Swagelok end-caps fitted with ¼ inch PTFE ferrules ready for transport.

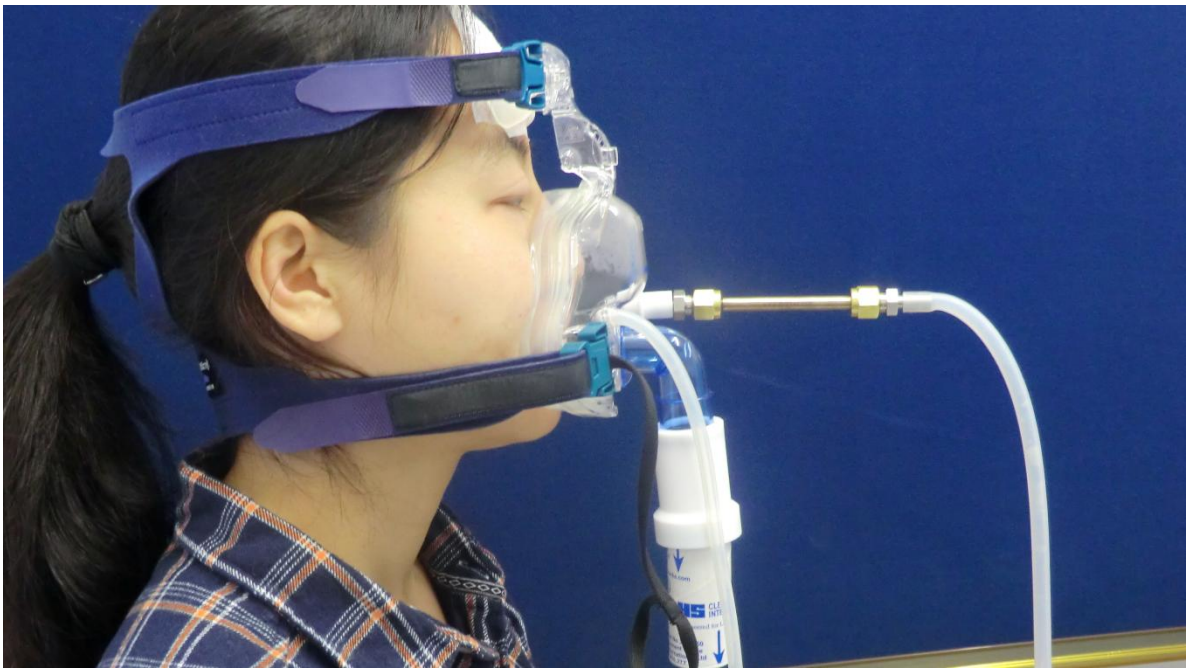


Figure 2-9 Photograph of the adaptive breath sampler in use.

2.4 Instrument design

The design of the thermal desorption – gas chromatography – mass spectrometry/differential mobility spectrometry system (TD-GC-MS/DMS) is integral to this work. Many of the volunteers involved in this work were too ill to request more than a

single sample from. This means that amount of data/information obtained from each sample must be maximised. The rationale and specification of analytical systems and methods used is necessarily a compromise between sensitivity and resolution. Data processing

Once samples have been collected and analysed it is necessary to classify and characterise the data produced. This is perhaps the most important and challenging part of the work-flow. Vast amounts of data can be produced rapidly, but with no methods to process and analyse the information such endeavours have little intrinsic value. The methods used in this research were: Baseline correction, normalisation, deconvolution, principle component analysis and discriminant function analysis. Further a new processing method for the analysis of dispersion field programmed differential mobility spectrometry was an important aspect of this research.

2.4.1 Thermal desorption

Thermal desorption (TD) is a technique commonly used to desorb volatile organic compounds from adsorbent traps such as those seen in Figure 2-4. The basic operation of a thermal desorption instrument is to heat the adsorbent trap to desorb the volatile organic compounds from the adsorbent surface and flush them from the trap with a flow of inert gas. The released volatiles are then concentrated in a cold-trap before injection, by flash-heating onto a chromatography column.

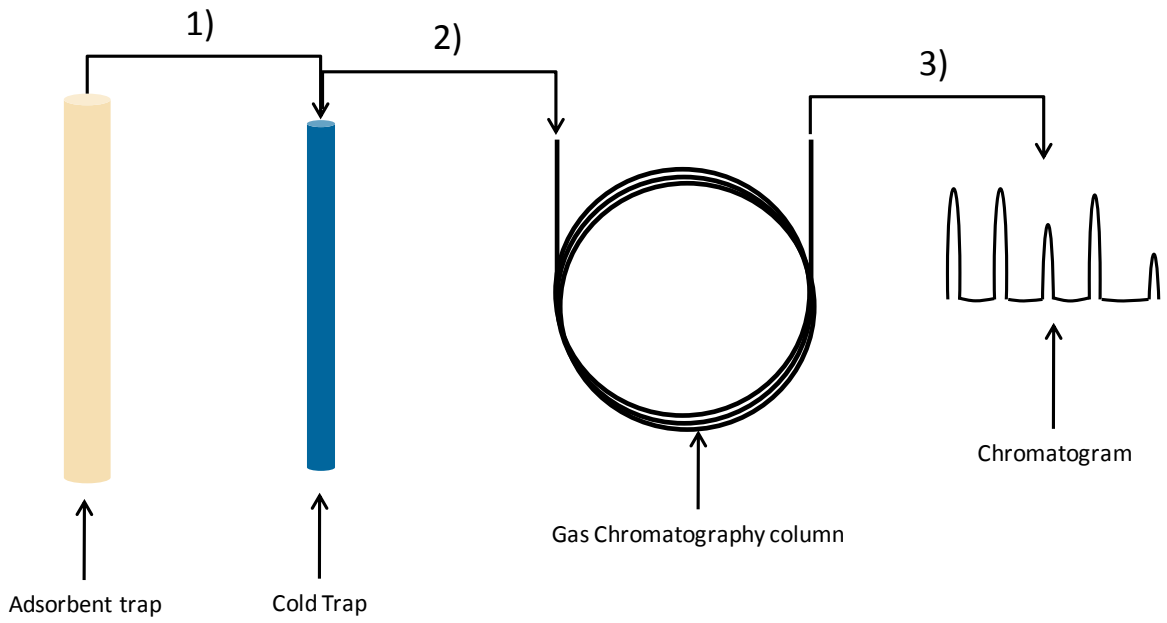


Figure 2-10 The flow of analytes through an analytical system. 1) Shows the analytes flowing from the adsorbent trap to the cold trap. 2) Shows the flows of analytes from the cold trap to the gas chromatography column and 3) shows the volatiles flowing from the gas chromatography column to the analyser to produce a chromatogram.

The cold trap contains a much smaller volume of adsorbent held at a lower temperature (usually -20 °C and 5 °C , depending on the adsorbent used). This results in the volatiles being re-trapped in the cold trap on this much smaller volume. The cold trap is then heated very rapidly resulting in the volatile organic compounds being injected as a much smaller volume, significantly improving sensitivity and chromatography. TD combined with cold trapping also has the benefit of enhancing the concentration of the sample. This can be shown by calculating the concentration enhancement (CE);

$$CE = \frac{D_c}{S_c}$$ **Equation 2-1**

Where D_c is the detected concentration and S_c is the sampled concentration. If it is assumed that the sampled volatile organic compounds are completely desorbed, then the concentration is proportional to the volume and CE can be re-written as;

$$CE = \frac{S_v}{D_v}$$ **Equation 2-2**

Where S_V is the sampled volume and D_V is the desorbed volume. With a sampled volume of approximately two and a half litres and a cold trap volume of approximately 100µL, the concentration enhancement is 2.5 x 10^4. This makes thermal desorption a very useful tool when analysing volatile organic compounds.

Thermal desorption systems can be easily interfaced with gas chromatography columns making them an excellent choice of injector when adsorbent traps are being used.

2.4.2 Gas chromatography

Gas chromatography is a technique for separating complex mixtures of volatile compounds. The volatiles partition between a stationary phase, in the case of gas chromatography this is a thin film of polymer, and a mobile phase, the carrier gas. It is the activity of the analyte in the two phases that gas chromatography exploits to separate the analytes. In practical terms gas chromatography provides a way of separating out a mix of volatile compounds, such as a breath sample, before they reach the detector. This makes it easier to identify particular biomarkers. The key terms used in gas chromatography are listed in Table 2-1 along with their definitions.

Table 2-1 The important terms in gas chromatography along with their definitions.

Term	Definition
Column oven	A programmable oven that can be used to control the temperature of the column contained inside. The Column oven has built in injectors (the injector type depends on the oven manufacturer) and mounting points for detectors. They also provide communication connections for any additional components such as external injectors and detectors, so they can be started/stopped by the column oven.
Temperature programme	A temperature programme is a pre-defined set of temperatures and times that dictate what temperature the column is at any point during the chromatographic run.
Chromatographic run	The chromatographic run is the time from when the sample is injected to when detection is finished.
Gas chromatography column	The gas chromatography column is a capillary column of a specific length and diameter. The inside surface of the capillary is lined with a liquid film stationary phase (which can be of different composition depending on the applications) that causes separation of the analytes as they pass through the column and partition between it and the mobile phase.
Mobile phase	The mobile phase is the gas which carries the analytes along the chromatography column. This gas is inert so as not to reach with the analytes and is usually helium.
Internal column diameter	The internal column diameter is the internal diameter of the chromatography column. A larger internal diameter increases the volume of the mobile phase and so changes the partitioning of the analytes.
Film	The film is the stationary phase. It is composed of a polymer that can be varied to interact more with specific analytes. Increasing the interaction between the film and the analytes increases the interaction between the two and so alters the partitioning of the analytes.
Film thickness	The film thickness is the thickness of the film applied to the inside of the chromatography column. Increasing this will increase the volume of the stationary phase and change the partitioning of the analytes.

Analysis of breath samples by gas chromatography raises four key challenges. All of these must be addressed as best they can when deciding upon the method to be used. These challenges are:

- A wide range of unknown volatiles to be analysed,
- The high water content of breath samples,
- The stability of the chromatography system and,
- The limitations imposed upon the system by the detectors.

The selection of the column has a large influence on the separation of the analytes. There are a wide range of gas chromatography columns available for analysing volatiles. They differ in length, diameter, stationary phase film composition and stationary phase film thickness and all of these things have an influence over the separation of the analytes. As such a wide range of analytes is being investigated a "general purpose" column that is not targeted towards any particular analytes should be used. The column selected was 30 m long with an internal diameter of 0.25 mm. The stationary phase selected was a 95 % methyl:95 % phenyl with a film thickness of 0.25 µm.

The high water content of the breath samples can cause problems for the analysis. The stationary phase of a gas chromatography column is degraded by water and high concentrations can cause damage which results in changes to the chromatography and can reduce the separating power of the column. To reduce the amount of water as much as is feasible when analysing breath it is important to select the compatible adsorbents for your adsorbent traps, described in 2.3.4. Also, using a hydrophobic cold trap in the thermal desorption unit and maintaining its' temperature above zero degrees centigrade can help reduce the amount of water injected into the chromatography column. Careful monitoring of the intensity of the background contaminant levels in system blanks can help determine if the column is degrading.

The stability of the chromatography system must be maintained over the course of the experiments. Deterioration of the column phase, changes in the carrier gas pressure and

fluctuations in the temperature programme can all cause problems. Monitoring the condition of the chromatography column using system blanks can again indicate if it is reaching the end of its' useful lifespan. Another method of correcting for changes in the chromatography system is to use retention indices instead of retention time when classifying a compound. A retention index ladder can be created by running standards through the chromatography column at regular intervals. The retention time of each peak in the system can then be converted into a factor of the retention times of the standards, known as a retention index. As the column ages and the chromatographic properties change over time the retention time of the standards will change. Using the most recent standards retention time values to re-register the data converts them to retention indices and will make every run compatible.

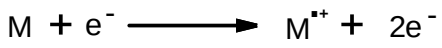
Limitations may be imposed upon the chromatographic separation by the detectors. In this work the DMS used dispersion field programming [109] which limited the analytical run time to one hour. Using a multi-step temperature gradient can help to focus separation on particular areas of the chromatogram and help overcome some of the limitations imposed upon the separation.

2.4.3 Mass spectrometry

Mass spectrometry as a technique began in nineteen hundred and six, when J.J. Thomson discovered the electron and determined its mass-to-charge (m/z) ratio which led to him constructing the first mass spectrometer in nineteen hundred and twelve [110]. There have been many advances in the technique since this first instrument leading to many different mass analysers, which have varying properties, but all still separate ions based on their m/z ratio. There has also been a large number of ion sources developed which are used depending on the type of analysis being carried out. There is a vast amount of literature on mass spectrometry, with many textbooks giving broad overviews or focussing on particular techniques or areas of application, so the descriptions in this section will focus only on the techniques used in this work.

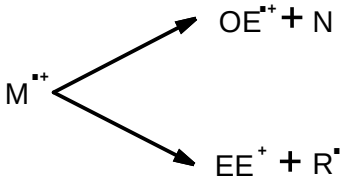
2.4.3.1 Principles of mass spectrometry

To measure the mass of a molecule in a mass spectrometer, the molecule must be ionised. The simplest method of this is electron ionisation;



Equation 2-3

Where M is a molecule and e is an electron. Ions of this form are likely to undergo fragmentation, because it is a radical, which can progress in two ways;



Equation 2-4

Where $M^{\bullet+}$ is the molecular ion, $OE^{\bullet+}$ is an odd ion fragment, $R^{\bullet}$ is a radical, EE^{+} is an even ion fragment and N is a new molecule. The two new ions formed have different properties and can themselves' undergo fragmentation. A mass spectrometer will separate these ions based on their m/z ratio, detect them and give and output based on their abundance. Multiply charged ions can also occur in a mass spectrometer and they too are separated based on their m/z ratio.

The mass analyser separates out the different ions based on their mass to charge ratio (m/z). There are six different types of basic analyser but in general terms they all do this same basic operation. All mass analysers make use of either an electric or magnetic field. This is why an analyte must be ionised be it enters the mass analyser. Neutral analytes are not affected by the electric/magnetic field and thus cannot be analysed. The mass analysers' ability to distinguish between different ions is called the analysers resolution or resolving power. The resolution of a mass analyser can be defined by the following equation:

$$Resolution = M/\Delta M$$

Equation 2-5

Where M is the m/z and ΔM is the full width at half maximum. In practice the higher the resolution the more able a mass analyser is to distinguish between two different ions. Resolution is limited by the instrumentation and its' mode of action. For example a time-of-flight instrument has a theoretically infinite mass range. This is because the ions are accelerated along a flight tube with the same energy and as such smaller ions will have a greater velocity and so will fly along the tube faster than larger ions. As separation is a function of the tube length, in an infinitely long tube all ions would eventually reach the detector. Of course in practice this is not possible. Limitations are imposed on the tube length as it must be maintained at constant temperature and under a high vacuum as changes in length due to heat expansion or contraction and any collisions within the tube would result in the ions reaching the detector at a different than expected point and produce a wrong result for the m/z.

A benefit of certain analysers, such as a quadrupole ion trap, is their ability to perform MS^n experiments. An MS^n experiment is when ions are selected and fragmented, detected and then a fragment is re-selected and fragmented, detected and so on. These experiments are particularly useful for protein and peptide analysis but they are also useful for determining between isomeric compounds. They can also provide more information about a particular analyte when the resolution of the instrument is not capable of providing a monoisotopic mass. A monoisotopic mass is a mass that relates to a particular molecular formula composed of particular isotopes of each element. Instruments capable of measuring monoisotopic masses are usually limited to a certain point where it is able to distinguish between the different isotopic peaks for a single compound. As the mass of a compound increases the amount of different isotopes of the elements expected also increases and this results in extra peaks in the mass spectrum separated by very small amounts. When an instrument is no longer capable of distinguishing between two monoisotopic peaks they will be combined and 'centroided'. Centroiding is when the average of the two peaks is displayed as a single bar in the mass spectrum. Fragmentation of analytes in instruments that are either incapable of achieving monoisotopic mass resolution or when the mass is above the monoisotopic mass limit of the instrument can provide structural information about the analyte and allow for identification without the need for an exact mass.

2.4.3.2 Quadrupole ion trap mass analyser

The quadrupole ion trap mass analyser was first described by Paul and Steinwedel in 1953 [111] and was modified by Stafford *et. al.* [112] of Finnigan company. A quadrupole ion trap analyser is made up of a circular electrode with two ellipsoid caps above and below it, shown in Figure 2-11.

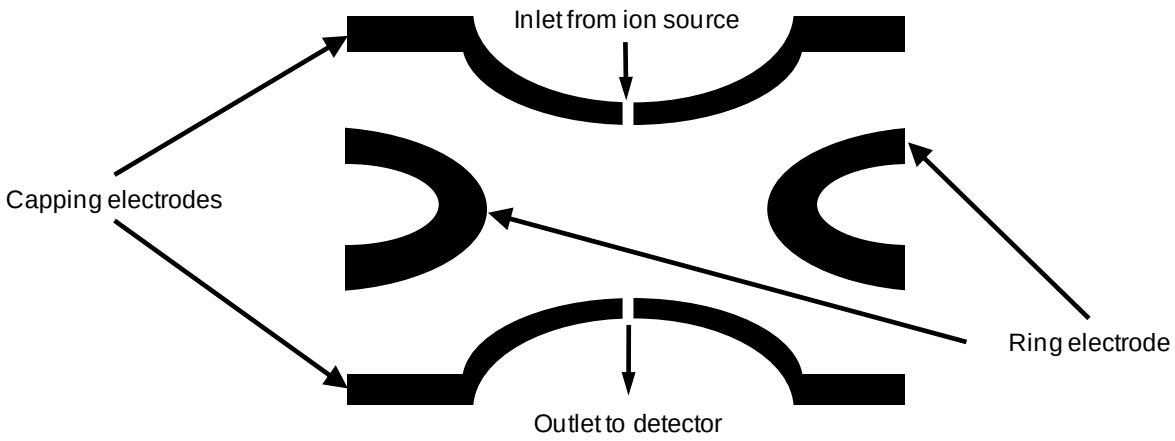


Figure 2-11 Schematic view of a quadrupole ion trap showing the ring electrode and the capping electrodes above and below it.

The motion of the ions within the trap is three dimensional and controlled by the potentials applied to the electrodes. The three dimensions can be labelled as x, y, and z with the trap having a radius of r_0, as shown in Figure 2-12.

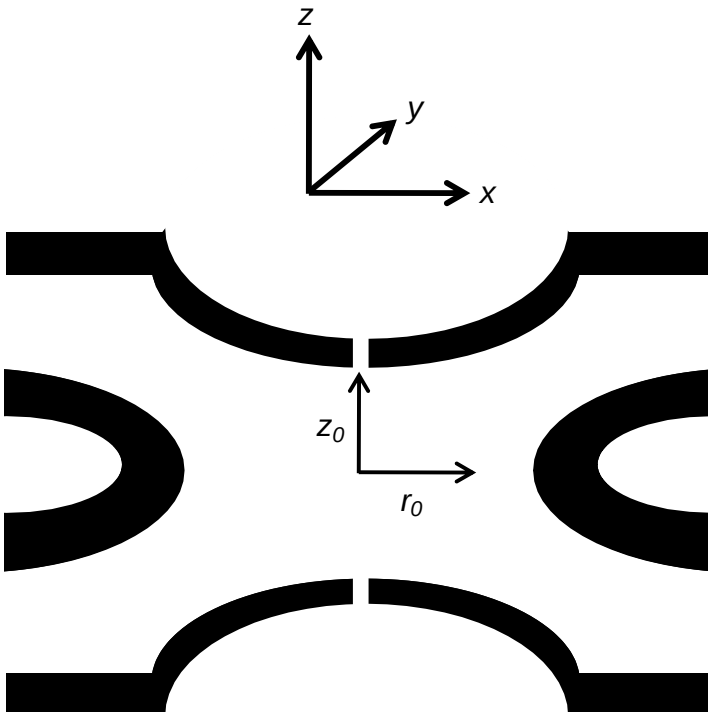


Figure 2-12 Direction of the *x*,*y*,*z* coordinates and the distances z_0 and r_0 within the trap.

As long as the coordinates of x and y never reach or exceed r_0, and the coordinate z never exceeds z_0, then the ions will have a stable trajectory.

The ions within a quadrupole ion trap are held there by the RF frequencies applied to the two electrodes andcan be ejected from the trap in two different ways, ejection at the stability limit and resonant ejection. Ejecting at the stability limit is limited by the dimensions of the trap and the maximum value of V. Larger masses require a larger value of V to reach instability, this can be reduced by reducing the dimensions of the ion trap, but this would make the trap more prone to arcing. Resonant ejection can extend the mass range, as it is possible to calculate a particular frequency at which an ion, within the trap, will oscillate. By applying this frequency along the z-axis, an ions' oscillations can be increased to a point where it is ejected from the trap along this axis. The main consideration with resonant ejection is fragmentation within the trap. If an ion within the trap fragments, it is possible that it will be ejected at a frequency that corresponds to a different ion, if V is being increased to scan across a range of ions. This will result in an artefact peak in the mass spectrum which corresponds to a different m/z than the fragments actual value.

The resolution of an ion trap is determined by the scan speed. This means that the resolution can be increased greatly by reducing the scan speed of the ion trap, and resolutions of up to 30,000 have been reported [113], but this can greatly reduce the sensitivity of the instrument due to the increased scan time and is not typical.

The main consideration when using a quadrupole ion trap is the 'space charge effect'. If too many ions are present in the trap at any one time, ions oscillating closer to the electrodes, 'shield' ions oscillating closer to the centre of the trap. This changes the frequency at which the shielded ions are ejected and introduces mass errors. One way of taking this effect into account is to monitor the total ion count (TIC). When this reaches a maximum value, specified by the instrument manufacturer, space charge effects are likely to be seen. An introduction to quadrupole ion trap mass spectrometry by R. March provides a good introduction to this topic [114].

2.4.3.3 Considerations when using ion trap mass spectrometry for breath analysis

The use of ion trap mass spectrometry for breath analysis has some challenges. Space charging, as previously described in 2.4.3.2, can cause shifting of m/z peaks, poor chromatography can result in complex fragmentation patterns that are difficult to separate, the large orders of magnitude difference between the analytes can result in a loss of sensitivity and the high levels of water in the breath samples can cause degradation of the column and introduce large background artefact peaks.

Poor chromatography, caused by increases in background noise and artefact peaks, can cause problems with accurately assigning an identification for analytes. Overlap of chromatographic peaks means that fragmentation patterns also overlap and these must be separated out so an analyte can be identified. The use of chemical ionisation (CI) may help with this but this would introduce more problems as not all analytes will ionise using CI, such as aliphatic hydrocarbons. This would also add the problem of ion suppression, with ions that are more likely to pick-up a charge ionising preferentially.

The large difference in orders of magnitude of analytes in the samples can cause two problems:

1. Smaller peaks may be lost under larger peaks if they co-elute

2. In the presence of space charging and combined with the challenges of deconvolution of complex chromatography the shifts in m/z can cause automated deconvolution programmes to assign m/z values as two different analytes when in fact they are due to a single analyte.

The high levels of water in breath samples can increase the effect of several problems all ready described as well as creating some of their own. Water degrades the chromatography column resulting in causing large artefact peaks of siloxanes and high levels of background noise from increased column bleed, the gradual break-down of the column, in the mass spectra. This can increase the difficulty associated with deconvolution as well as increasing the likelihood of space-charging. As the compounds released by column damage tend to be higher weight compounds they can increase the need for trap cleaning and maintenance. It can be estimated that a two and a half litre breath sample contains approximately 0.16g of water, more water than adsorbent used.

2.4.4 Differential mobility spectrometry (DMS)

Differential mobility spectrometry (DMS) is a technique for separating gas phase ions. It can be coupled to a gas chromatography system [115] or used as an isolated sensor for detecting airborne volatiles [116]. As an analytical system it is small, low-cost and requires few laboratory services, compared to mass spectrometers. As an on-site analysis technique these features of DMS are highly desirable, this makes it an interesting candidate for evaluation for on-site breath analysis. It has already been demonstrated for breath analysis [117, 118] and research in this area is growing.

The history of DMS begins, like mass spectrometry, with J.J. Thomson. His work on the mobility of ions in an electric field, first published in his book in 1903 [119], lead to him receiving the Nobel Prize for physics in 1906. His work relates to the linear form of ion mobility spectrometry (IMS). As a technique, IMS was primarily developed by E. W. McDaniel during the nineteen fifties and nineteen sixties, where he developed a drift tube and measured the mobility of H^{3+} and H^{+} ions in room air [120]. This method of IMS

measures the time taken for ions to drift along a tube against a buffer gas, under an applied low electric field. The equation for movement of ions in an electric field comes from Thomson's work:

$$V_d = KE$$ **Equation 2-6**

Where V_d is the drift velocity of the ion in cm s^{-1}, K is the ions mobility constant in $cm^2\ V^{-1}\ s^{-1}$ and E is the applied field strength in V cm^{-1}. As can be seen from this equation, for different ions to separate they must have a different drift velocity, and therefore must have a different mobility constant. The mobility constant of an ion is controlled by several factors and can be approximated by the newly reconsidered Mason-Schamp equation [121]:

$$K = \frac{1}{8}\left(\frac{M+2m}{M+m}\right)\sqrt{\frac{2\pi}{\mu kT}}\frac{q}{N\Omega_D}$$ **Equation 2-7**

Where M is the mass of the analyte, m is the mass of the gas, q is the unit charge on the ion, N is the number density of the drift gas, μ is reduced mass of the ion, k is the Boltzmann constant, T is the temperature of the drift tube and Ω_D is cross-sectional area of the ion. This means that, as long as everything is kept constant, the time taken for an ion to reach the detector will be dependent on only the reduced mass and its collision cross-sectional area.

A linear time-of-flight IMS drift tube is often composed of a series of field defining electrodes conductive rings separated by insulators which form an air-tight drift region, an ionisation source (typically ^{63}Ni), separated from the drift region by an shutter grid, and a Faraday plate; used to detect any ions as leave the drift region. A schematic of an IMS tube can be seen in Figure 2-13. Neutral analyte molecules are introduced into the ionisation source and are held there by the shutter grid. Once ionised a packet of ions can be injected into the drift region by removing power to the shutter grid. This power is then restored to trap the next packet of ions. Without this shutter grid the ions would enter the drift region in a continuous stream and it would not be possible to separate them. A series of voltages applied across the field defining electrodes produces a linear voltage gradient across the field defining electrodes which accelerates ions through the drift region. A drift gas flows in

the opposite direction to the motion of the ions under the electric current. The action of the voltage on the ions combined with the collisions with the drift gas causes ions of different sizes to attain limiting velocities and so reach the detector at different times, separating different ion species.

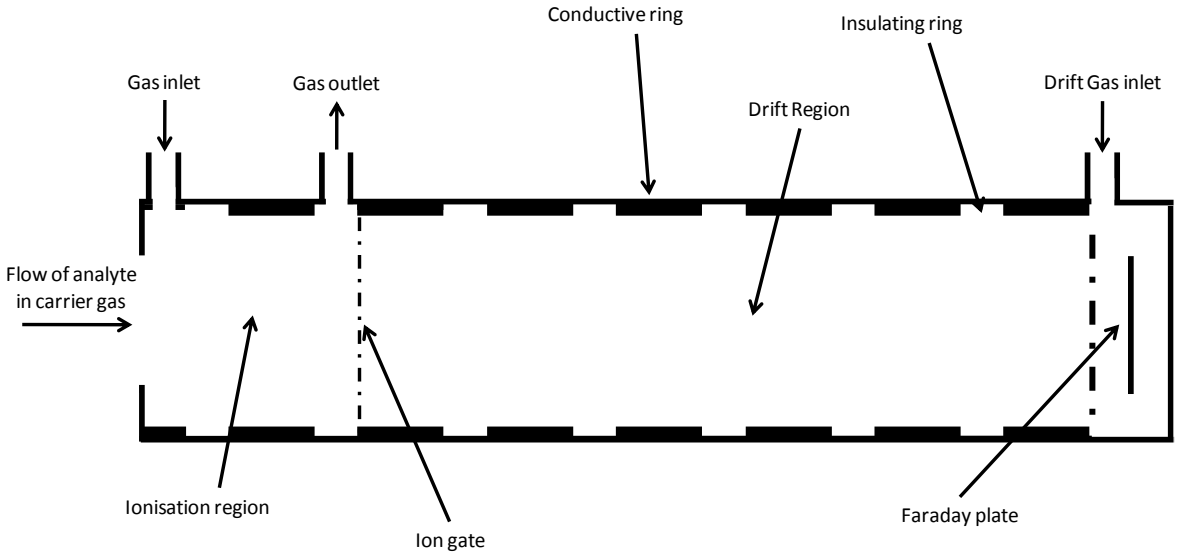


Figure 2-13 Schematic of an IMS device showing the ion source, the drift tube constructed of conductive and insulating rings, the ion gate, the Faraday plate and the inlets and outlets for the reactant gas, the drift gas and the carrier gas.

DMS as a technique was first demonstrated by Buryakov *et. al.* in 1993 [122]. A schematic of a DMS instrument is shown in Figure 2-14. By applying a high frequency voltage to one plate, the ions' trajectory is changed and moves towards the grounded plate. Application of a compensation voltage to the plates to counter the drift enables ions to be selectively filtered through the analyser and hence to the detector measured.

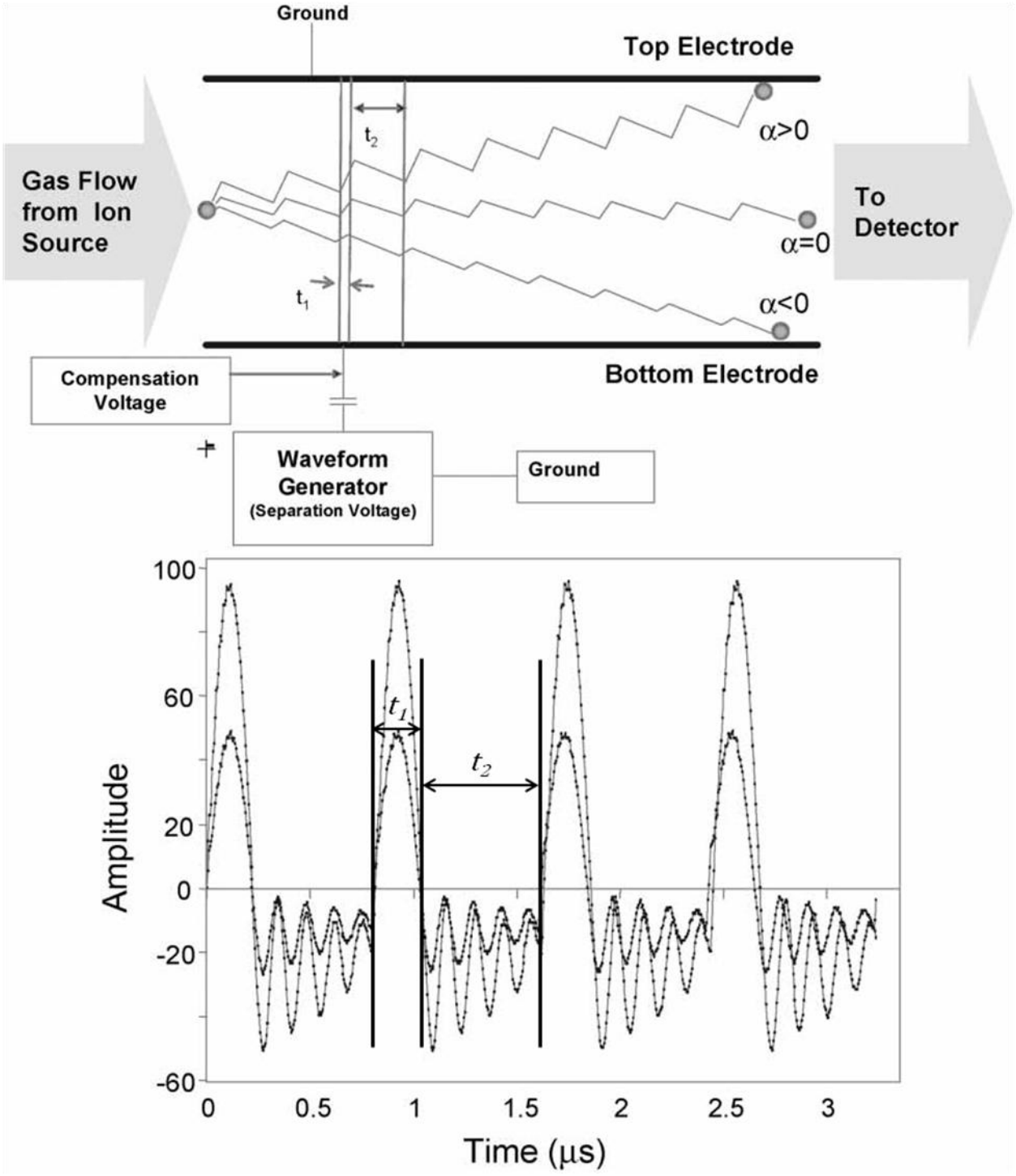


Figure 2-14 Schematic of a DMS system, showing the parallel plates which form the drift region , the waveform generator which provides the voltage to the bottom plate to separate the ions and the applied compensation voltage. Below this are two asymmetric waveforms with different semi-periods labelled as t_1 and t_2 for two different levels of voltage. Adapted from [123].

From the nineteen eighty eight book by Mason and McDaniel [124] the simplified form of the mobility constant *(K)* according to electric field *(E)* when density of the drift gas *(N)* is constant can be given as:

$$K_E = K_0[1 + \alpha_E] \qquad \textbf{Equation 2-8}$$

Where K_E is the field dependant mobility at electric field strength E, K_0 is the mobility constant at zero electric field and α_E is the function for the mobility constant at electric field strength E at constant number density of drift gas. Buryakov *et. al.* [122] described the separation of a mixture of ions using the following equations:

$$\frac{1}{T}\int_n^T E\,(t)dt = \langle E(t)\rangle = 0 \qquad \textbf{Equation 2-9}$$

$$\langle E^{2n+1}(t)\rangle \neq 0$$

where T is the field oscillation period and n is an integer not less than 1 describes a curve with period $T = t_1 + t_2$ where the maximum field strength (E_{max}) during t_1, the positive semi-period is much greater than the minimum field strength (E_{min}) during the negative semi period t_2. If t_1 is much less than t_2 this causes the ions to oscillate, for a short period, perpendicular to their flow in the drift gas during period t_1 according to E_{max}, E_{min} and α_E. This means that if E_{max} and E_{min} are constant, ions will be displaced perpendicular to their flow according to their α_E values. By creating an electric field between the plates according to:

$$E(t) - E_s(t) + E_c = E_s f(t) + E_c \qquad \textbf{Equation 2-10}$$

where E_s satisfies the conditions of Equation 2-9, $f(t)$ is the function which describes the form of the field, $E_s = max\ E_s(t)$ and E_c is a constant voltage known as the compensation voltage. Specific selection of E_c, E_s and $F(t)$ make it possible to direct specific ions towards the detector, and variation of E_c with E_s makes it possible to successively compensate for all of the ion species present, as shown in Figure 2-15.

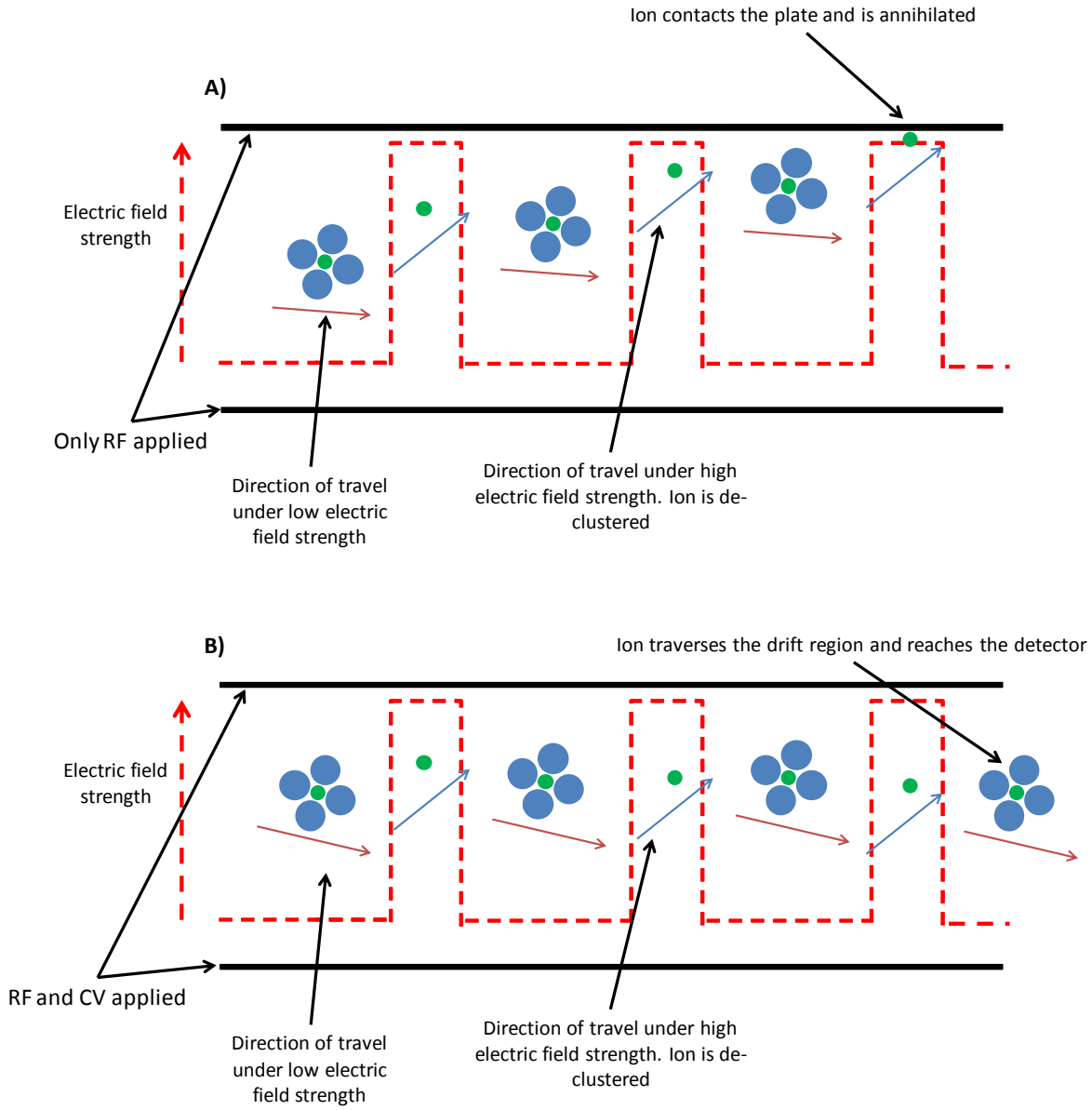


Figure 2-15 How applying a compensation voltage, E_c, can allow an ion to traverse the drift region. A) shows the motion of the ion without the application of a compensation voltage and B) shows the motion of the ion with a compensation voltage.

By varying the compensation voltage across a set range a final output from the instrument of signal intensity Vs. E_c can be created as seen in Figure 2-16.

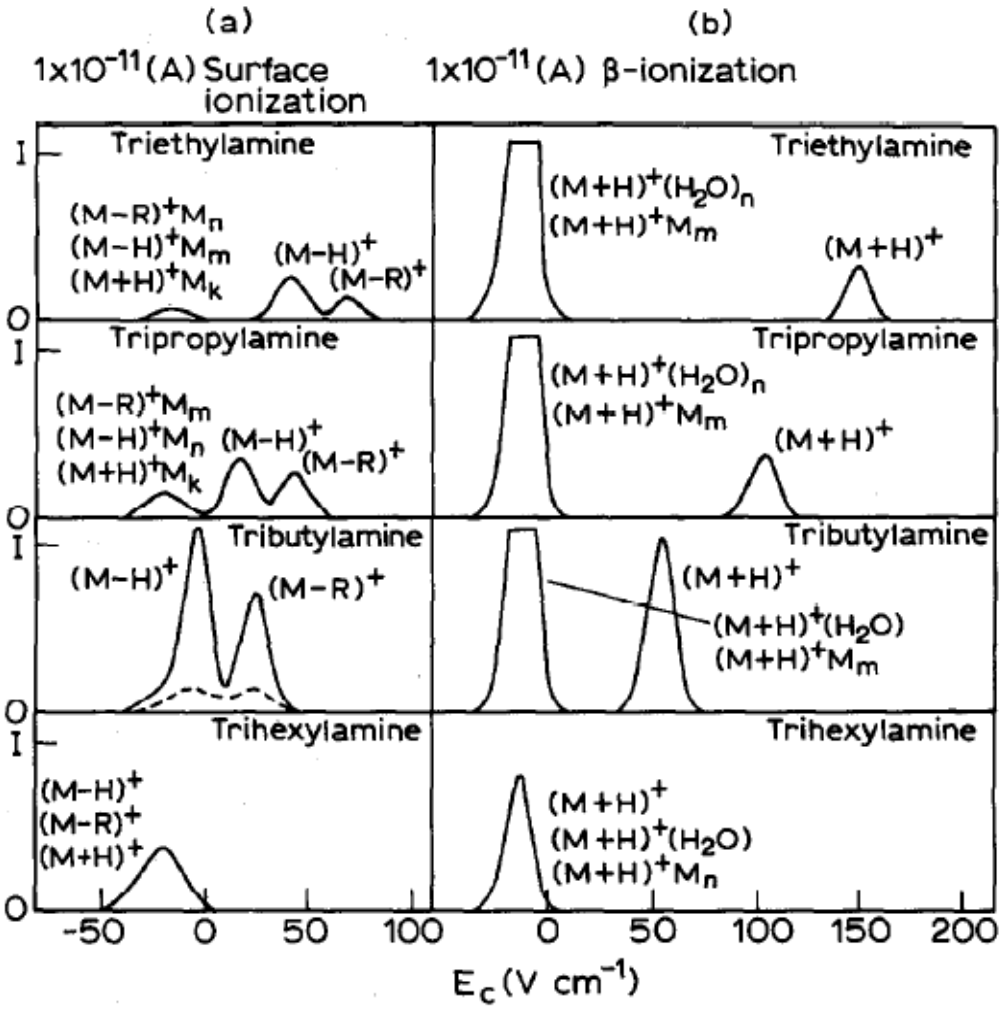


Figure 2-16 Compensation voltage versus intensity plots taken from Rasulev and co-workers [122] showing the drift spectra of substances produced by (a) surface ionisation and (b) β-ionisation.

This technology forms the basis of the Sionex Micro DMx chip, the DMS system used in this work. As a detection technique DMS exploits the ions change in mobility at different field strength to separate out complex mixtures of ions. In this way it differs from traditional IMS. This allows DMS to be used for continuous analysis methods without the need for ion gates that are present in IMS systems. This is because at a particular high field strength the point at which different ions reach the detector is dictated by the compensation voltage rather than drift time.

As with mass spectrometry, DMS requires the molecules to be ionised before analysis. The method used in this work, and one of the more common methods is atmospheric chemical

ionisation (APCI) via a β-particle emitting ^{63}Ni source. As with all APCI sources, several reactions can occur, depending on what molecules are present in the source at the time of ionisation. The main benefits of a ^{63}Ni source are:

- it produce positive and negative ions which can be measured simultaneously in a DMS system,
- is stable and long lasting,
- small, compact and easy to integrate into the inlet of an instrument and,
- requires no external power.

Reactant ions are generated by ^{63}Ni ionisation of a gas stream, usually nitrogen or air. This description focuses on nitrogen, as this was used in this work. The following series of reactions describe the formation of the common positive reactant ions [125]:

$$63Ni \rightarrow e^- \quad \textbf{Equation 2-11}$$

$$N_2 + e^- \rightarrow N_2^+ + 2e^- \quad \textbf{Equation 2-12}$$

$$N_2^+ + N_2 \rightarrow N_4^+ \quad \textbf{Equation 2-13}$$

$$N_4^+ + H_2O \rightarrow 2N_2 + H_2O^+ \quad \textbf{Equation 2-14}$$

At this point it is important to note that water is required for all but one species of reactant ion, as N_2^+ and N_4^+ are high energy ions and will likely not survive as ions long enough to traverse the drift region [126]. The reactions proceed as follows:

$$H_2O^+ + H_2O \rightarrow H^+(H_2O) + OH \quad \textbf{Equation 2-15}$$

$$H^+(H_2O) + nH_2O \rightarrow H^+(H_2O)_{n+1} \quad \textbf{Equation 2-16}$$

The only species of reactant ion that does not require the presence of water to form is:

$$N_2^+ + NO \rightarrow NO^+ + N_2$$ **Equation 2-17**

The final reactant ion formed is:

$$H^+(H_2O) + NH_3 \rightarrow NH_4^+ + H_2O$$ **Equation 2-18**

However it should be noted that both the nitrogen containing reactant ions were shown to cluster with water when they were identified by Karasek and co-workers [127], with the three reactant ions being $H^+.(H_2O)_n$, $NO^+.(H_2O)_n$ and $NH_4^+.(H_2O)_n$, with the value of n being dependant on the temperature. It is the reactant ions.

The reactant ions will form product ions with the analytes introduced, mainly of the forms:

Protonated monomer -

$$M + H^+(H_2O)_n \rightarrow MH^+ + (H_2O)_n$$ **Equation 2-19**

Hydrated protonated monomer -

$$MH^+ + (H_2O)_n \rightarrow MH^+(H_2O)_n$$ **Equation 2-20**

Proton-bound dimer -

$$MH^+ + M \rightarrow M_2H^+$$ **Equation 2-21**

Hydrated proton-bound dimer -

$$M_2H^+ + (H_2O)_n \rightarrow M_2H^+(H_2O)_n$$ **Equation 2-22**

As well as adduct formation, charge-exchange ionisation also occurs, particularly with aromatic compounds. This involves a charge passing from an ion to a neutral molecule –

$$A^+ + B \rightarrow A + B^+$$ **Equation 2-23**

The maximum concentration of any ions formed is governed by the concentration of the reactant ions and any observed intensity due to product ions shows a corresponding decrease in intensity due to reactant ions.

In the negative ion mode the formation of ions differs greatly from that of the positive ion mode. With clean gas systems, such as pure nitrogen, the negative charge carriers are the electrons. In air there are several reactant ions, identified by Spangler and Collins [128] mainly as O_2^-, CO_4^- and $(H_2O)O_2^-$.

The ion clusters that are formed, which are dependent on: the amount of water present, the temperature of the instrument and the concentration of the analyte separate under the high field portion of the waveform. This is the effect that DMS exploits to separate out ions. However, as the size of the analyte molecule increases this effect becomes less pronounced, if the high field maximum remains the same, reducing the resolution of the instrument. While not much can be done about this if samples are directly injected to the instrument, combining DMS with front end gas chromatography means that it is possible to slowly increase the maximum field strength. As has been discussed, chromatography allows for separation of complex mixtures with a tendency for larger molecules to reach the DMS system later than smaller molecules. By programming the dispersion field to increase during the chromatographic separation it is possible to increase the analytical space to maximise the potential of the DMS system [109].

2.4.4.1 DMS as a detector for lung disease

For this work a dual detector approach was decided on using both a DMS and MS. There are two reasons this approach was decided upon:

- To extract the maximum amount of information possible from the breath samples and,
- To test the utility of a DMS as a point of care instrument.

The most important element in the design of the experimental protocol for collecting breath samples from people with serious lung diseases is their welfare. The time available for

sampling must be kept to a minimum and this limits the number of samples available to ensure adequate machine replicates running dual detectors from a single sample was the operational compromise sought.

DMS is a strong candidate for development as a point of care instrument. DMS systems are small, robust, inexpensive to make, require few services and may be made easy to operate. By including in parallel with the mass spectrometry data evaluation of the DMS for breath samples was possible.

One of the main considerations for using DMS as a point of care instrument is its' ability to detect the compounds identified as biomarkers of disease. A DMS is capable of detecting volatiles as long as they can be ionised and understanding the ion chemistry in the ^{63}Ni source is vital.

2.4.5 Ion chemistry in differential mobility spectrometry ^{63}Ni source for identified biomarkers of lung cancer

^{63}Ni ionisation chemistry is selective towards analytes with compatible ionisation behaviours, normally with proton affinities greater than that for the reactant ion peak, hydrated proton in this study. It is important to understand this ion chemistry and whether possible biomarkers for lung cancer will be ionised and detected in the DMS. Table 2-2 to Table 2-6 contain lists of potential biomarkers for lung cancer identified in the literature and whether they will be ionised by the ^{63}Ni source. They have been split into compound classes, aromatics, alkanes, ketones, alcohols and other compounds, with each compound class in a different table. Chen *et. al.* [129] identified 1-butanol and 3-hydroxy-2-butanone as markers of lung cancer. The authors point out that at temperatures over 30 degrees centigrade 3-hydroxy-2-butanone will be oxidised to 2,3-butanedione. The commonly accepted core temperature of the human body is 37.0 degrees centigrade as determined by Carl Reinhold August Wunderlich. While revisions have been made to this over time they are small and with this in mind it is questionable whether 3-hydroxy-2-butanone would survive to be sampled based on the authors' conclusions. For this reason it is not included in Table 2-4, 1-butanol is included as a possible marker. For an analyte to be ionised by the ^{63}Ni source it must be able to accept a charge. Compounds classes such as alcohols, aldehydes and

ketones will be ionised along with benzene and its' derivatives and cyclic hydrocarbons, as shown by G. A. Eiceman et. al. [123] Analytes that would not be ionised are compounds that are unlikely to form clusters such as saturated and unsaturated hydrocarbons.

Table 2-2 Aromatic compounds observed in human breath samples and the headspace of cell culture lines reported in the literautre. Adapted from [130] to include whether the compound will ionise in a ^{63}Ni source, whether it is a sensible I.D. for a biological marker and any comments about its' known uses. The *in vitro* studies are shaded grey.

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible I.D. for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
Benzene	x		x		x	x					Y	Y	
1,2,4-trimethylbenzene	x		x								Y	Y	Petrol additive
1,4-dimethylbenzene	x										Y	Y	
1-methylethenylbenzene	x										Y	Y	
1,1-(1,2-cyclobutanediyl)bis-benzene		x									Y	Y	Plasticiser
1,1-[1-(ethylthio)propylidene]bis-benzene		x									Y	Y	
1,1-ethylidenebis[4-ethyl-]benzene		x									Y	Y	
Ethylbenzene			x				x			x	Y	Y	
Trimethylbenzene			x								Y	Y	
Propylbenzene	x			x							Y	Y	
1-Methyl-4-(1-methylethyl)-benzene							x				Y	Y	
1-Methyl-2-(1-methylethyl)-benzene							x				Y	Y	
1-Methyl-3-(1-methylethyl)-benzene							x				Y	Y	
1-ethyl-4-(2-methylpropyl)-benzene										x	Y	Y	
(1,1-dymethylbutyl)-benzene										x	Y	Y	
1,3bis-(1-methylethyl)-benzene										x	Y	Y	
1,1-dimethylbutylbenzene										x	Y	Y	

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible I.D. for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
1,2,4-trimethyl-5-(1methylethyl)-benzene										x	Y	Y	
1-(1,1-dimethylethyl)-3-ethylbenzene										x	Y	Y	
1,2-benzenedicarboxaldehyde										x	Y	Y	
a` -hydroxybenzeneacetonitrile										x	Y	Y	
1,3,5-trimethyl-2-propylbenzene										x	Y	Y	
1-ethyl-3-(1-methylethyl)-benzene										x	Y	Y	
1,2-bis(1-methylethyl)-benzene										x	Y	Y	
(1,1-dimethylpropyl)-benzene										x	Y	Y	
Styrene	x		x	x			x			x	Y	Y	
Toluene			x			x	x			x	Y	Y	
Xylenes (o-, -p, etc…)			x				x			x	Y	Y	
Phenol										x	Y	Y	
2,6-Bis(1,1-dimethylethyl)-4-methylmethylcarbamate pieno							x				Y	Y	
1H-Indene, 2,3-dihydro-4-methyl		x									Y	Unlikely	petrochemical compound
Camphor		x									yes	Unlikely	Fragrance and food additive, used as a pasticiser, moth repellant

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible I.D. for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
Benzophenone		x									Y	Possible	
Butylated hydroxytoluene (BHT)							x				Y	Possible	Antioxidant found in food, pharmaceuticals etc.

Table 2-2 give the details of all the aromatic compounds observed in human breath and lung cancer cell culture lines as reported in the literature [34, 36-39, 87, 131-134]. Many of the different compounds in Table 2-2 are isomers of each other. As all of the studies use electron ionisation sources it would be difficult to assign a specific isomer to a mass spectrum, as their fragmentation patterns would be very similar. So while 35 different compound identifications have been made this number may be lower. DMS would provide separation in the compensation field axis of isomers so would go some way to identifying what isomers are present if it were run in parallel with a mass spectrometer. All of the compounds in this table would be ionised by the 63Ni source and so are suitable for analysis in the DMS. Several of the compounds are found in exogenous sources that may influence the analysis such as 1,2,4-trimethylbenzene which is a petrol additive and BHT which is used as an antioxidant. Camphor has been labelled as an unlikely biomarker as it is used in many products that the participants may come into contact with. 2,3-dihydro-4-methyl-1H-Indene is also labelled as unlikely as it is extracted from oil by the petrochemical industry.

Table 2-3 Alkyl compounds observed in human breath samples and the headspace of cell culture lines reported in the literature. Adapted from [130] to include whether the compound will ionise in a ^{63}Ni source, whether it is a sensible I.D. for a biological marker and any comments about its' known uses. The *in vitro* studies are shaded grey.

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
2,2,4,6,6-pentamethylheptane	x										N	Yes	
2-methylheptane	x										N	Yes	
methylcyclopentane	x										Y	Yes	
1-methyl-2-pentylcyclopropane,	x										Y	Yes	
trichlorofluoromethane,	x										Y	No	CFC-11
Cyclohexane	x										Y	Yes	
3-methylnonane	x										N	Yes	
3-methyloctane	x										N	Yes	
2,4-dimethylheptane	x										N	Yes	
1,1,2-trichloro-1,2,2-trifluoroethane		x									Y	No	CFC-113
2-methoxy-2-methyl-propane		x									Y	unlikely	Additive in petrol
Decane	x		x	x							N	Yes	
Undecane	x			x				x		x	N	Yes	
Pentane			x			x			x		N	Yes	
Heptane			x								N	Yes	
Octane			x								N	Yes	
Hexane									x		N	Yes	

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
1,2,4,5-Tetroxane, 3,3,6,6-tetraphenyl-		x									Y	Yes	
2-Methylpentane			x						x		N	Yes	
Pentamethylheptane			x								N	Yes	
Methylcyclopentane				x							Y	Yes	
3-Methylhexane							x				N	Yes	
3-Ethylpentane							x				N	Yes	
2,3,4-Trimethylhexane							x				N	Yes	
2,4-Dimethylheptane							x				N	Yes	
4,7-Dimethylundecane							x				N	Yes	
2,6,6-Trimethyloctane							x				N	Yes	
3,3-Dimethylpentane							x				N	Yes	
3,3-Dimethylhexane							x				N	Yes	
2-Methylhexane							x				N	Yes	
3-Ethylhexane							x				N	Yes	
2,2,3-Trimethylhexane							x				N	Yes	
Ethylidenecyclopropane							x				Y	Yes	
4-Methyloctane							x				N	Yes	
2,3,4-Trimethylpentane							x				N	Yes	
2,3-Dimethylhexane							x				N	Yes	
Isobutane						x					N	Yes	

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
2,3-Dimethylbutane									x		N	Yes	
3-Methylpentane									x		N	Yes	
Methylcyclopentane									x		Y	Yes	
2,3,4-Trimethylpentane									x		N	Yes	
2,3,3-Trimethylpentane									x		N	Yes	
4-Methylheptane									x		N	Yes	
2,4-Dimethylheptane									x		N	Yes	
2,3-Dimethylheptane									x		N	Yes	
4-Methyloctane									x		N	Yes	
EthylOxirane									x		Y	Yes	Carcinogen, mutagen
3-ethyl-5-methyl-1-propylcyclohexane									x		Y	Yes	
1,5-diethyl-2,3-dimethylcyclohexane									x		Y	Yes	
Cyclopropylphenylmethane									x		Y	Yes	
1-methyl-3-ethyladamantane									x		N	Unlikely	Pharmaceuticals, Petrol
4,7-dimethylundecane									x		N	Yes	

Table 2-3 give the details of all the alkyl compounds observed in human breath and lung cancer cell culture lines as reported in the literature [34, 36-39, 87, 131-134]. As with Table 2-2 there are several compounds in this table that are isomers of each other and it is unlikely that accurate isomeric identification of these analytes could be made using just mass spectrometry. This means that again the number of unique compounds identified may be less than is currently reported. There are two compounds listed as not being suitable identifications for biological markers in this table. They are trichlorofluoro-methane and 1,1,2-trichloro-1,2,2,-trifluoro-ethane which are both CFCs'. 2-methoxy-2-methyl-propane and 1-methyl-3-ethyladamantane are listed as unlikely. These both occur in petrol and 1-methyl-3ethyladamantane is an additive in pharmaceuticals. This means that both of these compounds are too likely to be exogenous and so can't be considered.

Very few of these compounds would be ionised in a ^{63}Ni source and this is a gap in the current technology. Aliphatic hydrocarbons are unlikely to pick up a charge and form clusters and this is a problem for analysis with the DMS. They will be ionised in a mass spectrometer however and this will allow for identification using the TD-GC-MS/DMS system described. If any aliphatic hydrocarbons were identified as markers for lung cancer then, if DMS is to be used as a point of care detector, work would have to be done on the ionisation of hydrocarbons, possibly using a different method of ionisation.

Table 2-4 Ketones observed in human breath samples and the headspace of cell culture lines reported in the literature. Adapted from [130] to include whether the compound will ionise in a ^{63}Ni source, whether it is a sensible I.D. for a biological marker and any comments about its' known uses. The *in vitro* studies are shaded grey.

Compound	Reference Observed in										Ionisable with 63Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
2,3-Hexanedione		x									Yes	Yes	Used as food additive
3-Hexanone, 2-methyl-		x									yes	Yes	
Bicyclo[2.2.1]heptan-2-one, 1,7,7-, (Camphor)		x									yes	No	Fragrance and food additive, used as a pasticiser, moth repellant
trimethyl-,(1S)-a` Isomethyl ionone		x										No	Fragrance additive
2,2,7,7-Tetramethyltricyclo[6.2.1.0(1,6)]undec-4-en-3-one		x										Yes	
2,5-Cyclohexadien-1-one, 2,6-bis(1,1-dimethylethyl)-4ethlylidene		x										Yes	
2,4,6-Tris(1,1-dimethyl-ethyl)-4-methylcyclohexa-2,5-dien-1-one							x					Yes	
2-Butanone							x		x			Yes	
2-Methyl-4,6-octadiyn-3-one							x					Yes	Volatile seen in onions eur.

					Jour. Plant. Path. 110: 371-377, 2004
3-Ethyl-3-methyl-2-pentanone		x		Yes	
6,10-Dimethyl-5,9-dodecadien-2-one		x		Yes	Found in Gentiana genus of plants biochem. Sys. Ecol. 33, 9, sept 2005, 938-947
Acetone	x			Yes	
Ethanone, 2-hydroxy-1-phenyl (HPE)			x	Yes	
2-(3-Methylbuta-1,3-dienyl)cyclohexanone			x	Yes	

Table 2-4 gives the details of all the ketones observed in human breath and lung cancer cell culture lines as reported in the literature [34, 36-39, 87, 131-134]. As can be seen from the comments column, several of the ketones are found in products and food stuffs. As these compounds have an exogenous source, identifying them as biomarkers of lung cancer is made more difficult as it can be difficult narrowing down exogenous sources. This is why the breath sampling system used in this work is designed to minimise exogenous sources of volatile compounds by using a full-face mask. Ketones will ionise in a ^{63}Ni source so they would be seen in the DMS.

Table 2-5 Aldehydes observed in human breath samples and the headspace of cell culture lines reported in the literature. Adapted from [130] to include whether the compound will ionise in a ^{63}Ni source, whether it is a sensible I.D. for a biological marker and any comments about its' known uses. The *in vitro* studies are shaded grey.

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
Hepatanal	x										Y	Y	
Hexanal													Perfume additive, used in
	x			x					x		Y	Y	flavour industry
Formaldehyde					x						Y	Y	
Heptanal				x					x		Y	Y	Perfume additive
Acetaldehyde													Occurs in coffee bread and ripe
									x		Y	Y	fruit
2-Methylpropanal									x		Y	Y	
3-Methylbutanal													Product of 3-methyl-butanal
									x		Y	Y	reductase
2-Methylbutanal									x		Y	Y	
Benzaldehyde										x	Y	Y	metabolite of amygdalin
3-ethylbenzaldehyde										x	Y	Y	

Table 2-5 gives the details of all the aldehydes observed in human breath and lung cancer cell culture lines as reported in the literature [34, 36-39, 87, 131-134]. As with the other compound tables, there are several aldehydes that can be identified as possibly coming from an exogenous source. Two of the aldehydes in Table 2-5 have been identified as products of metabolism. This highlights that it is perfectly reasonable for aldehydes to be proposed as biomarkers and the volatility of smaller aldehydes means that they can be sampled by breath analysis. Many aldehydes appear in foods naturally and this means that it again may be difficult to eliminate exogenous sources and careful collection and analysis of associated metadata along with using sampling methods that minimise the chance of exogenous interference are the best way to do this.

Table 2-6 Other compounds observed in human breath samples and the headspace of cell culture lines reported in the literature. Adapted from [130] to include whether the compound will ionise in a ^{63}Ni source, whether it is a sensible I.D. for a biological marker and any comments about its' known uses

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
p-menth-1-en-8-ol (alpha Terpineol)		x											Perfum additive, flavour in Lasang Souchong
Isopropyl alcohol		x									Y	Y	Cleaning solvent
5-Isopropenyl-2-methyl-7-oxabicyclo[4.1.0]heptan-2-ol		x											
2,2,4-Trimethyl-1,3-pentanediol diisobutyrate		x										N	Plasticiser
9,10-Anthracenediol, 2-ethyl-		x										N	possible metabolite of a 9,10-anthracenedione based drug
4-Penten-2-ol		x									Y	Y	Possible component of perfume
Isopropanol					x	x							
Ethanol						x	x						
2-Ethyl-1-hexanol							x		x	x		N	causes diarrhoea and vomiting, nausea and headache
2-Ethyl-4-methyl-1-pentanol							x			x		Y	

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
2-Propyl-1-pentanol							x					Y	
Methanol						x						Y	
1-propanol									x			Y	
Terpineol												Yes	Fragrance additive
1-hexene	x			x								Y	
1-heptene	x											Y	
1-(methylthio)-(E)-1-propene		x									Y	Y	
5,5-Dimethyl-1,3-hexadiene		x									N	Y	
Isoprene	x		x	x		x		x		x	N	Y	
1,3,5-Cycloheptatriene							x				N	N	
1,3,5,7-Cyclooctatetraene							x					Possible	Isolated from Fungi
Bicyclo[4.2.0]octa-1,3,5-triene							x				N	Y	
1,3-Pentadiene							x				N	Y	
2,4-Dimethyl-1-heptene									x		N	Y	
Naphthalene										x	Y	Y	
Azulene										x	N	Y	
1H-Indene, 1-methylene										x	Y	Y	petrochemical compound
2-propenylidenecyclobutene										x	N	Y	
1,2,3,4-tetrahydronaphthalene										x	N	Y	
Decahydro-2,6-dimethylnaphthalene										x	N	Y	

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
Decahydro-1,5-dimethyl Naphthalene										x	N	Y	
Decahydro-2,3-dimethylnaphthalene										x	N	Y	
Tricyclo[3.3.2.0(2,8)]deca-3,6-diene										x	N	Y	
4-ethoxy-ethylbenzoate		x									Y	Y	
Bicyclo[3.2.2]nonane-1,5-dicarboxylate		x									Y	Y	
2,2,4-trimethyl-3-carboxyisoproplisobutylpentoate		x									Y	Y	
2-methyl-,1-(1,1-dimethylethyl)-2-methyl-1,3-propanediylpropanoate		x									Y	Y	
Dimethylether							x				Y	Y	Propellant and fuel
2-[(2-ethoxy-3,4-dimethyl-2-cyclohexen-1-ylidene)methyl]-		x									Y	Y	
Vinylfuran										x	Y	Y	
Tetrahydrofuran										x	Y	Possible	
1,2,3,4-tetrahydro-9-		x									Y	Possible	Scintillator

Compound	Reference Observed in										Ionisable with ^{63}Ni source	Sensible ID for biological marker	Comments
	[36]	[87]	[131]	[132]	[133]	[134]	[34]	[37]	[38]	[39]			
propylanthracene													
2,3-dihydro-2-methylbenzofuran										x	Y	Y	
Hydrazine-carboxamide							x				Y	N	
Methylhydrazine							x				Y	No	Used as rocket propellant
Hydrazine							x				Y	No	rocket fuel
Carbonicdihydrazide							x				Y	N	
Dimethylsulfide						x					Y	Y	indicator of bacterial infection in brewing, found in foods, e.g. Beetroot, cabbage maize
Carbondisulfide						x					Y	Y	
n-Butylacetate									x		Y	possible	Flavour additive in cheese, ice cream, candy etc.
(p-hydroxyphenyl)phosphonic acid										x	Y	Y	
Methylbenzoate										x	Y	Y	
Benzoate hydrazide										x	Y	Y	
Acetic acid										x	Y	Y	
Nitrous oxide										x	Y	possible	

Table 2-6 gives the details of all the other compounds observed in human breath and lung cancer cell culture lines as reported in the literature [34, 36-39, 87, 131-134] that do not fall directly under one of the other table contents. These compounds again have been suggested as biomarkers for lung cancer. Some of the compounds that have been suggested are not suitable identifications for biomarkers. Hydrazine and the four hydrazine containing compounds are inorganic compounds so would not appear in the human body. 2-ethyl-1-hexanol causes diarrhoea, vomiting, nausea and headaches so it is unlikely that would be a biomarker for lung cancer. 2-ethyl-9,10-anthracenediol is possibly a metabolite of an anthrancenedione anti-cancer drug.

As can be seen from Table 2-2 to Table 2-6, many of the compounds identified as possible biomarkers for lung cancer also occur from other, exogenous sources. While steps can be taken to eliminate exogenous compounds as much as possible, this cannot be completely guaranteed. Careful collection and analysis of metadata is important to make sure that once a biomarker is identified possible exogenous sources can be eliminated.

2.5 Clinical studies

The clinical studies strand involves several different steps for it to be completed successfully. Taken from Figure 2-1 these are;

- Ethical approval,
- Study design,
- Study implementation and,
- Sample collection and transportation.

2.5.1 Ethical approval

Ethics is a philosophical discipline that seeks to discuss morality and associated ideas (for example good and evil, or right and wrong). In the context of this work, medical ethics are the values that are evaluated when deciding if an experiment on human subjects should go ahead. The Hippocratic Oath, believed to be written by Hippocrates around the 5th Century BC, is a document that outlines a set of ethical standards to be adhered to. There are two well known documents that have shaped modern medical ethics, the Nuremburg Code and the Helsinki Declaration.

Table 2-7 A modern version of the Hippocratic Oath [135]

1. I swear to fulfill, to the best of my ability and judgment, this covenant:
2. I will respect the hard-won scientific gains of those physicians in whose steps I walk, and gladly share such knowledge as is mine with those who are to follow.
3. I will apply, for the benefit of the sick, all measures [that] are required, avoiding those twin traps of overtreatment and therapeutic nihilism.
4. I will remember that there is art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife or the chemist's drug.
5. I will not be ashamed to say "I know not," nor will I fail to call in my colleagues when the skills of another are needed for a patient's recovery.
6. I will respect the privacy of my patients, for their problems are not disclosed to me that

the world may know. Most especially must I tread with care in matters of life and death. If it is given to me to save a life, all thanks. But it may also be within my power to take a life; this awesome responsibility must be faced with great humbleness and awareness of my own frailty. Above all, I must not play at God.

7. I will remember that I do not treat a fever chart, a cancerous growth, but a sick human being, whose illness may affect the person's family and economic stability. My responsibility includes these related problems, if I am to care adequately for the sick.
8. I will prevent disease whenever I can, for prevention is preferable to cure.
9. I will remember that I remain a member of society, with special obligations to all my fellow human beings, those sound of mind and body as well as the infirm.
10. If I do not violate this oath, may I enjoy life and art, respected while I live and remembered with affection thereafter. May I always act so as to preserve the finest traditions of my calling and may I long experience the joy of healing those who seek my help.

The Nuremburg code was written as a direct result of the Doctors' trial. The Doctors' trial was the trial of twenty three people, twenty two men and one woman, for war crimes. All of the defendants were accused of being involved in Nazi human experimentation. The defendants had argued that what they did could not be described as illegal as there was no law declaring what was legal and illegal. In defence of medical research, Dr. Leo Alexander had earlier outlined six points in defence of medical research in his capacity as Chief Medical Advisor to Telford Taylor the Chief of Counsel for War Crimes during the U.S. Nuremburg Military Tribunals. The six points outlined by Leo Alexander and a further four points were adopted by the trial verdict and formed the "Nuremburg Code". The ten points of the Nuremburg Code are;

1. The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an

understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonable to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment.

The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.

2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.

3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.

4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.

5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.

6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.

7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.

8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.

9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.

10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

The Helsinki Declaration was issued by the World Medical Association (WMA) as a set of principles for medical research involving human subjects. It was first issued in nineteen sixty four and has been amended six times, most recently amended in 2008 by the General Assembly of the WMA. Paragraph one of the Declaration states "The World Medical Association (WMA) has developed the Declaration of Helsinki as a statement of ethical principles for medical research involving human subjects, including research on identifiable human material and data." The Declaration of Helsinki is probably the best known policy statement of the WMA.

While the Declaration of Helsinki is not a set of legal guidelines it is often used to define what research is ethical and can be used when a case for research is presented to an ethics review board. The World Health Organization (WHO) state in their guidelines for contributors to their publications that "The World Health Organization publishes the results of research involving human studies only if it has been conducted in full accordance with ethical principles, including the provisions of the World Medical Association Declaration of Helsinki (as amended by the 59th General Assembly, Seoul, The Republic of Korea, October 2008" [136] and as such has been used in this work to help guide the ethics application.

Two European directives exist, the Clinical Trials Directive and the Good Clinical Practice (GCP) Directive that influenced the Law in the United Kingdom with respect to medical research. Clinical trials of medicinal products are controlled by the Clinical Trials Regulation, but this does not control other types of medical research that have no direct legal regulation. These trials, including the clinical study in this work, are covered by a complex set of common law, clinical governance and professional guidance and ethics applications are approved by the Regional Board of the National Health Service (NHS) Research Ethics Committee for where the research will be taking place and the Local Board of the NHS Research Ethics Committee at the NHS institution where the study will take place [137-139].

The ethics board will consider several different aspects of the work when reviewing an application for their opinion and will ultimately weigh up the "risk/benefit" of the research. This means that they will consider the potential risk to the subjects, which can be examined by looking at any supporting data for the application, and compare this to any potential benefits the research may provide. This does not mean that research with limited potential benefits will be instantly discarded or that research with little to no risk will be approved, each application will be considered on its' own merits. The key decision in this research was to investigate patients referred to the chest clinic and not specifically lung cancer sufferers. It was decided that this would be more likely to be granted ethics approval. This is because people with newly identified lung cancer will be under considerable stress and the treatment schedule can be very intensive. It was felt that it might cause undue stress to request that they participate in a study during this early period. By asking people to participate during their visit to the chest clinic no added stress would be incurred and there would be no disruption to their treatment schedule.

There are several stages to an ethics application that must be completed and approved before a clinical study can be commenced. In addition to the ethics applications researchers also need to satisfy health and security vetting criteria before they will be allowed to work with volunteers; the research passport. These processes must be completed before research with human participants can commence.

2.5.2 Study design

There are several different aspects that must be considered when designing a clinical study. The main challenge for researchers who are not working within the NHS is integrating into the NHS location where the research takes place and making sure this is possible is a large part of designing a successful clinical study. With this in mind the study should always be planned with the clinical lead. Access to volunteers and hospital sites is key to any clinical study and the clinical lead is best placed to provide this.

The study design should outline every aspect of the clinical study. These are:

- The aims of the research,
- The procedures involved in the research,
- The number of participants required for the study,
- The data analysis to be carried out and,
- The expected outcomes of the research.

All of these should be covered in the study documentation that is submitted to the Regional NHS ethics review board for approval. The study design allows all of the people involved in the research to be aware of what the basis for the clinical study is and how it should be carried out.

Clearly without ethics approval the research programme fails and the decisions about when and where to sample needed to reflect the clinical work-flow and possible impacts on individuals who may be gravely ill. It was decided that volunteers would be sought from individuals attending “the two week wait clinic”, rather than specify people with lung cancer. It would involve the taking of two breath samples per participant for twenty diseased participants and twenty healthy participants. Twenty participants in each group were decided on as this would give a wide range of biological variation, so any identified differences would be more likely to be markers of disease, but was a suitably small number of samples to collect in the time frame. These breath samples would be analysed at Loughborough University with a view to characterising the differences in breath samples

observed in the sampled cohorts. Identification of a suitable sampling site was the first and probably most important aspect of the design. To start with the sampling procedure must be:

- comfortable for the participants and;
- provide reproducible samples for an extended period.

See Section 2.3 for details on the breath sampler and technique used.

2.5.3 Study implementation

The implementation of the study is critical to the research. While the study will have been well-planned it needs to be a living thing and able to change if there should be a problem. With this in mind there should be a back-up plan for the different aspects of the study protocol. A second location for taking samples, different methods for recruiting participants (these would require ethical approval) should recruiting prove problematic, back-up sampling equipment in case there are problems with it and more than one person trained to use it should the operator be taken sick are all vital if the study is to proceed as planned.

A Gantt chart [140] is a useful document for following the planning and implementation of a study as is a flow chart that can provide quick reference to how each different aspect of the study interacts and these should be used by the study leader to maintain the smooth running of the work.

In general the study should be fairly straightforward to implement as it will follow the agreed and ethically approved programme. In the case of work such as that described here that is carried out between two sites constant communication between both sites is essential to ensure the study runs smoothly. Simple things such as making sure that the clinical researcher taking the samples in the clinic has sterile sampling kits of the right size for participants can prevent appreciable delays, particularly if sampling only occurs on a single day per week. Every effort should be made to avoid this.

2.5.4 Sample collection and transportation

Sample collection and transportation is the most important aspect of the study. Collection of the samples must be done under controlled conditions. Sampling in the same location, using the same air supply and sampling equipment is vital to maintain the consistency of sampling. The operator needs to make sure they are not contributing to any exogenous contamination of the samples or sampling kit. This means refraining from wearing perfumes and other scented cosmetics and using hand-creams and other products that may contain volatile compounds.

As the samples are volatile in nature all efforts should be made to ensure their fast transport to the laboratory for analysis. The volatile analytes are adsorbed onto the adsorbents in the adsorbent trap. To maintain this adsorption the samples should be transported at zero degrees centigrade to reduce the chance of more volatile analytes being lost.

2.6 Data processing and analysis methods

The main consideration when processing data is the avoidance of introducing bias to the analysis - something that is surprisingly easy to do, either intentionally or unintentionally, and this must be carefully monitored. While the obvious points where an analysis can be biased are easily dealt with, such as performing the same processing techniques to all the data sets, there can be more subtle effects that can go unnoticed and produce artefacts. For example: normalisation of the data set can remove relevant concentration effects between samples, while alignment of peaks can force peaks to align that are not actually equivalent. It is necessary and helpful to justify the use of each data processing step and describe fully the effect of each processing technique on the data.

Three different data processing methods were used:

- Baseline Correction,
- Normalisation,
- Euclidian distance transformation and,

2.6.1 Baseline correction

The noise in the analytical instruments signal is subject to day-to-day drift and this can result in a variation in different baseline values between analyses. To account for this a baseline correction on the data is performed subtracting the value of the minimum data point in the data set from every point. This means that every sample produced by a detector will have a comparable baseline. An example of this process is shown in Figure 2-17.

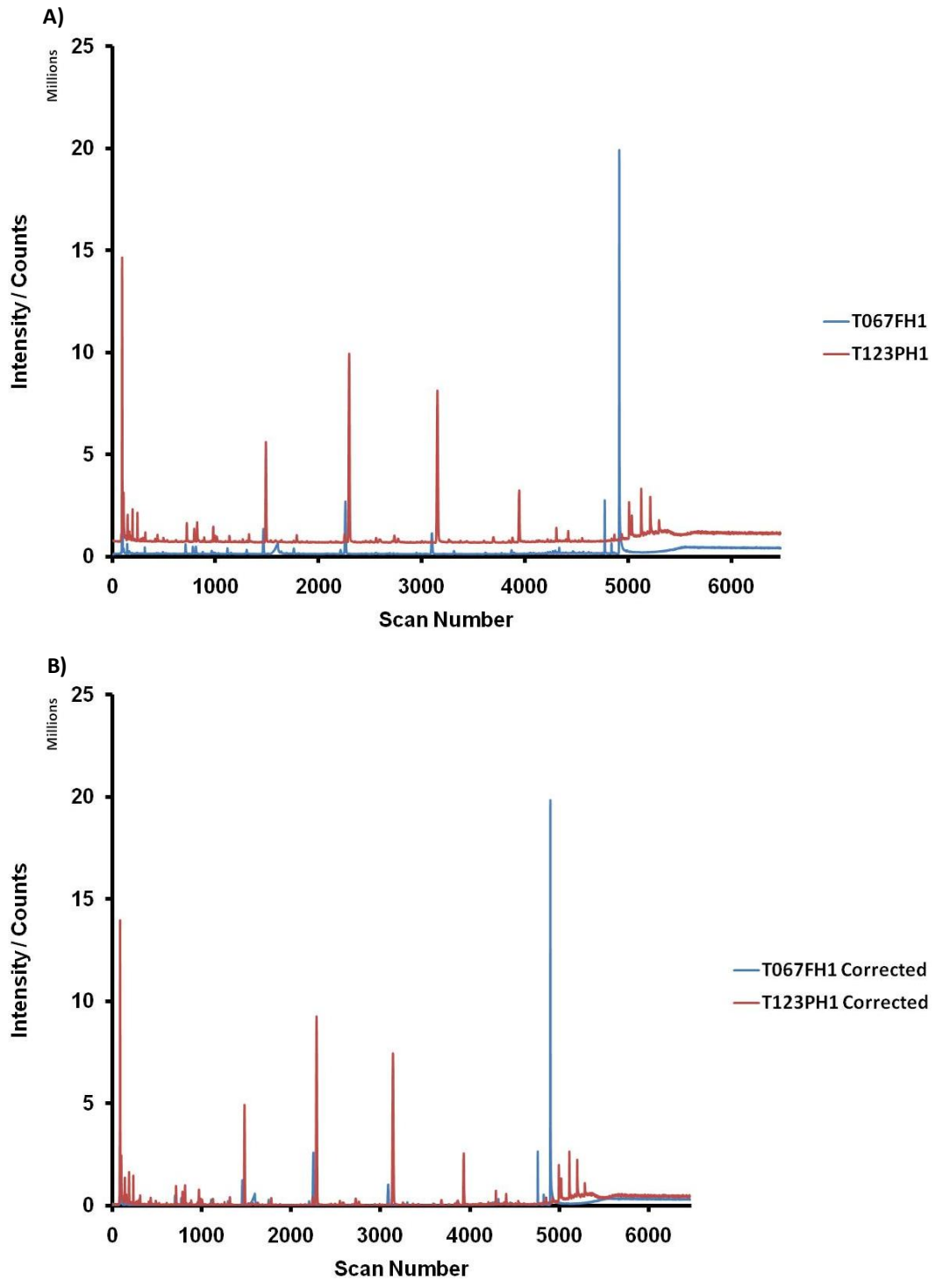


Figure 2-17 Two plots A) and B) of two breath samples before and after baseline correction. Plot A) shows the two breath samples plotted as the raw data produced by a Varian Ion Trap 4000 mass spectrometer. In this plot the baseline of T123PH1 has been artificially adjusted so the process is easier to visualise. Plot B) shows the breath samples after they have been baseline corrected. These two samples now have a comparable baseline.

2.6.2 Normalisation

Normalisation refers to rescaling the intensity of data against a specified range; most often on a scale zero to one. This can either be done for each sample or the sample set as a whole. This means that either each sample is scaled from zero to one individually or the sample set is viewed as a whole and largest and smallest values set at one and zero respectively, with all the other values adjusted accordingly.

The main consideration with normalisation is that concentration effects can be hidden both within and between samples.

2.6.3 Deconvolution

As analytes flow through the chromatography column they separate. If they completely separate then the peaks produced be the mass spectrometer will be due to one single analyte and can be identified. If they do not completely separate then this leads to convoluted peaks, peaks which overlap each other. If this happens they must be separated. This would usually be done based on the underlying mass spectrometry data.

2.7 Data analysis methods

2.7.1 Multivariate analysis

Multivariate analysis is the application of multivariate statistics, where more than one statistical variable is simultaneously observed. It is applicable to mass spectrometric data and differential mobility spectrometry data as both techniques measure a large number of variables in each run. There are many different multivariate statistical methods available for the investigation of variance in the data.

2.7.2 Principal Component Analysis (PCA)

Principal component analysis is a technique that reduces the dimensionality of a data set composed of a large number of interrelated variables while retaining as much as possible of the variation. It does this by transforming the data to a new set of uncorrelated variables; called principal components. Principal component analysis is an unsupervised method of data analysis that uses no prior knowledge of the samples.

Each principal component is calculated successively to account for the maximum amount of variance in the sample set. This means that the first component will account for the largest portion of the variance in the sample set, the second for the second largest portion and so on. Principal components are calculated by transforming the axes of the original plot so that the new axes lie along the maximum variance of the data. Figure 2-18 shows the first two principal components for the example data set shown. As can be seen in Figure 2-18, the first principal component lies in the direction of the maximum variance of the sample set, the second lies in the direction of the next highest variance and so on and so on.

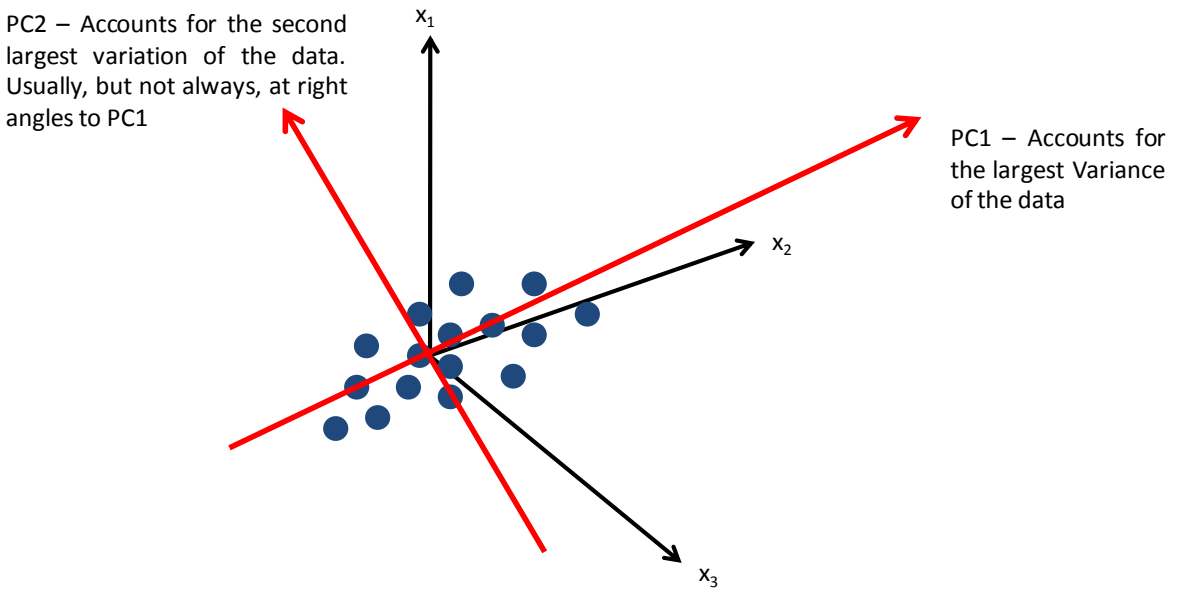


Figure 2-18 Example of the first two principal components projected onto a data set.

The scores for each principal component are the distance the point is from the origin after projecting the data points onto the principal component, as shown in Figure 2-19. This means that the points can now be plotted on a new set of axes according to the principal component scores to look for patterns in the data. Any observed patterns may be due to any number of factors and so it is necessary to fully investigate a data set.

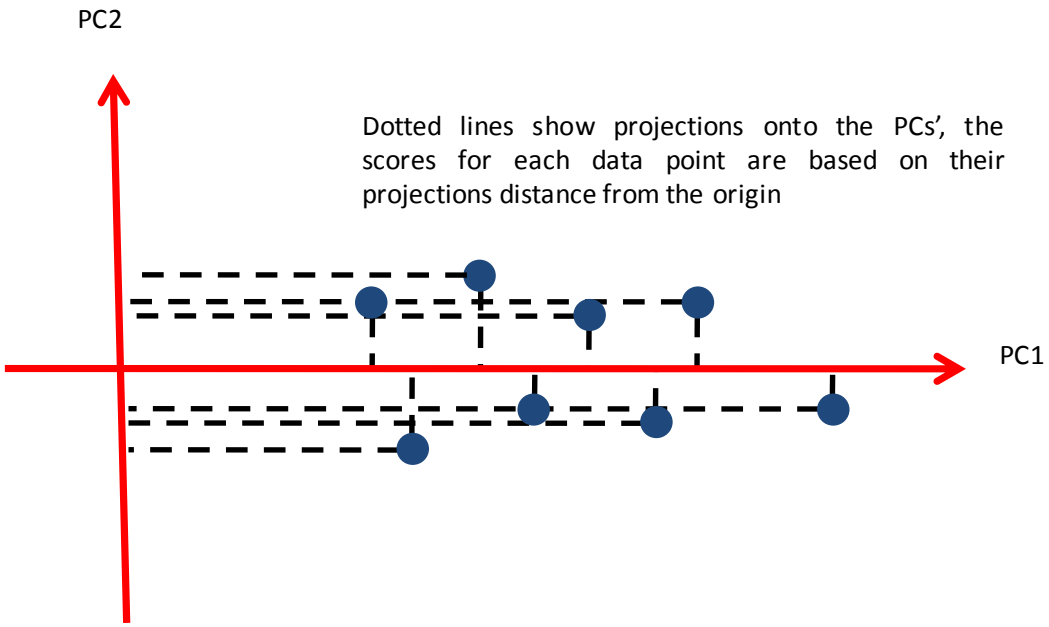


Figure 2-19 Example of how scores are calculated for each data point.

To fully investigate a data set the points can be plotted in terms of class. The class is the label applied to the points in the data set. This can include any meta-data as well as the initial sample classes, such as diseased and healthy or the date the samples were taken or analysed.

The number of principal components calculated is determined by the person doing the analysis. It is important to remember that the percentage of variance each component contains decreases for each successive component. This means that real differences between classes are most likely observed in the earlier principal components and it is possible that differences observed in the later principal components are due to artefacts for the analytical process and this must be considered when interpreting the results[141].

2.7.3 Discriminant Analysis (DA)

Discriminant analysis is a supervised method that uses a priori knowledge to classify samples to a specific group. It is broken down into linear discriminant analysis (LDA) where the grouping variable has two categories and multiple discriminant analysis where the grouping variable has more than two categories. Because it is supervised, discriminant analysis can test a prediction by observing if samples are assigned to the expected groups [142].

3 A dual channel thermal desorption gas chromatograph with differential mobility spectrometric and mass spectrometric detection.

3.1 Introduction

This work focuses on exploring the differences in breath between asymptomatic participants and participants with unambiguously defined lung disease. In doing this testing the utility of using differential mobility spectrometry (DMS) for the in-clinic analysis of breath was also an important task. In short, was it possible to establish a DMC-based method for detecting and identifying compounds found in breath samples.

The current "gold-standard" for analysing breath samples is gas chromatography - mass spectrometry (GC-MS) [35, 62, 85, 104, 129, 143-148]. This allows for separation of complex mixtures of volatile organic compounds (VOCs) and identification of the analytes. DMS cannot give identify unknown compounds without reference to known standards or comparison to mass spectral data. Pairing it with MS enables better characterisation of the DMS responses as well as the identification of compounds.

The challenge in this approach was to use both detectors for the analysis of every sample. There are advantages to running the detectors in series, such as the IMS-TOF instruments made by TOFWERKS [149], the SYNAPT series made by Waters [150] or the FAIMS adaption for Thermo mass spectrometers [151]. This approach was not available in this work as the DMS system employed, the Sionex S-VAC, cannot be combined in series with a Varian 4000 quadrupole ion trap mass spectrometer. An important additional consideration is that running the instruments in series would mean that the ionisation would be fixed on APCI, and the complementary information from electron impact ionisation would not be available.

A significant element of this study involved the modification of a commercial GC-MS to accept a DMS to run in parallel from a single sample and this chapter describes the modifications and protocols designed and undertaken to build a TD-GC-DMS/MS.

3.2 Instrumentation

3.2.1 Overview of instrument set-up

As a quick reference guide an overview of the system set-up is given. Consumable parts for connection of the separate components are given along with part numbers. The description is given from the sample injection point, the thermal desorber, to the detection point, the mass spectrometer and the differential mobility spectrometer.

1. A Markes International UNITY thermal desorber was connected by its transfer-line through an empty injection port on a Varian 3800 Gas Chromatography oven to a VSIS-2 inlet splitter (SGE part No. 123632) via a GVF/004 (SGE part No. 072663) ferrule.
2. Two Agilent J&W DB-5ms (I.D. 0.25 mm, 30 m, 0.5 µm Agilent part No. 122-5536) analytical columns were connected to the VSIS-2 inlet splitter via a GVF2/004 (SGE part No. 072662) ferrule.
3. Both of the columns were connected to a SilTite™ mini-union via an 85 per cent Vespel/15 per cent graphite 0.1-0.25 mm I.D. ferrule (SGE part No. 072696).
4. The DMS was connected to the SilTite™ mini-union via a length of deactivated fused silica capillary (SGE part No. 0624431), this length was determined as required to keep the SilTite™ mini-union inside the Varian 3800 gas chromatography oven. The capillary column was connected to the DMS using a 1/32” Valco adaptor ferrule (Thames Restek part No. 20137). The fused silica capillary was passed through the annular heat pipe in the DMS mount.
5. The MS was connected to the SilTite™ mini-union via a length of deactivated fused silica capillary (SGE part No. 0624431), this length was determined as required by the length of the transfer-line heater to the mass spectrometer. The fused silica capillary is connected to the mass spectrometer through the transfer-line via a 85 per cent Vespel/15 per cent graphite 0.1-0.25 mm I.D. ferrule (SGE part No. 072696).

3.2.2 Adsorbent trap conditioning

All adsorbent traps used in this work were conditioned before use. This was done at 280^{o}C in an adapted gas chromatography oven for at least 8 hours. During conditioning nitrogen was passed through the tube at 150 mL min^{-1}. After conditioning all tubes were analysed by TD-GC-DMS/MS to check they were blank before being shipped for samples collection.

3.2.3 Outline of the instrumentation system.

The instrument design enabled two detectors to be operated in parallel from the sample that was split onto two GC-columns immediately after injection; one column per detector, see Figure 3-1.

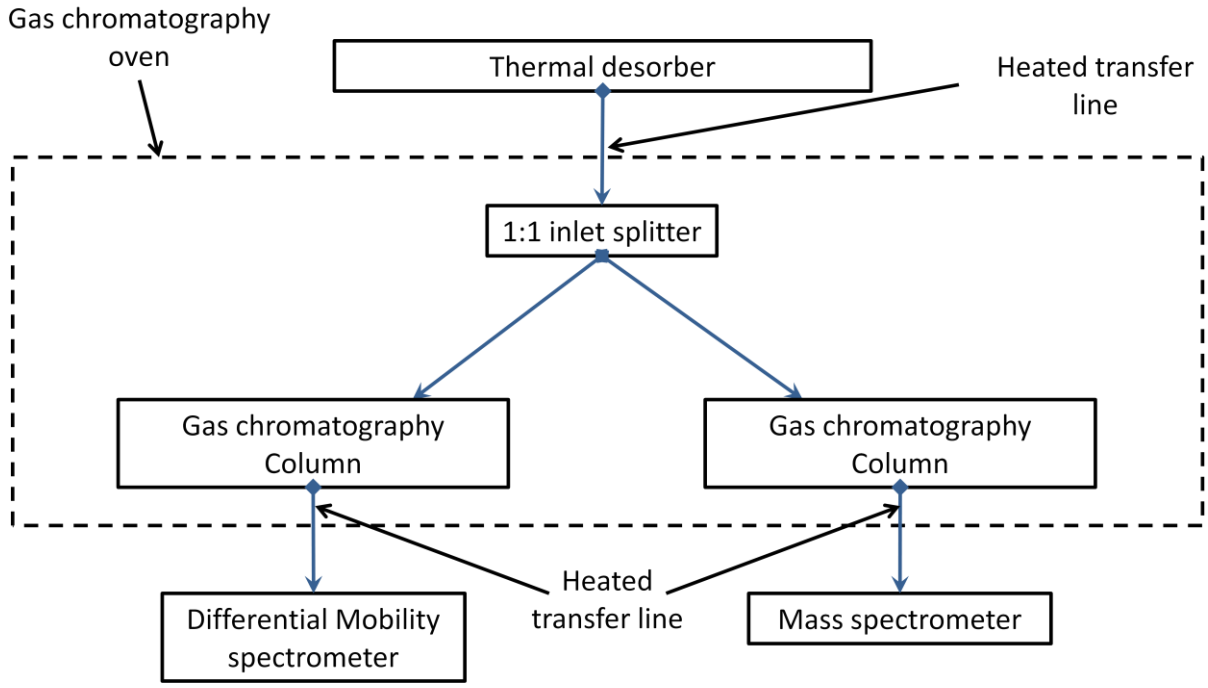


Figure 3-1 A flow diagram to show the flow of analyte through the parallel system. All transfer lines between each part of the system are heated.

This system was based on: Varian gas chromatograph (GC 3800); an ion trap mass spectrometer(Varian 4000) with a Markes International UNITY thermal desorber. A 1:1 splitter (SGE part No. 123632) was fitted to the end of the transfer line form the thermal desorption unit as it entered the gas chromatography oven. A panel was removed from the top of GC and a detector mounting port was modified to receive a heated transfer line and a Sionex S-VAC differential mobility spectrometer.

3.2.4 Differential mobility spectrometer mount

There is no specific mount or fitting to connect the Sionex S-VAC unit to the Varian GC-3800 so a mount was designed to connect these two units, Figure 3-2 - Figure 3-4.

The mount was designed for the detector interfaces on the top surface of the Varian GC-3800 and the Sionex S-VAC DMS system. The Varian GC-3800 has several threaded holes to allow for mounting detectors above the oven, and holes through to the oven to allow GC columns to pass through to the detectors. A plate was machined to fit across the top of the Varian Detector outlets and anchor at four of the threaded mounting holes. Two threaded holes were cut in the plate to allow mounting rods to be secured in the base plate. These mounting rods allow connection of the Sionex S-VAC DMS unit to the base plate. A 44.3 mm hole was cut to allow an annular heat pipe assembly to be mounted to the base plate.

The length of the mounting rods was set by the dimensions of the S-VAC unit and the annular heat pipe. The mounting rods were machined from stainless steel. A 3 mm diameter 13 mm threaded stainless steel rod was screwed into a 5 mm deep threaded hole and glued in place. Using a separate rod, rather than turning down the original rod, makes for a stronger mounting point. These threaded rods allow the S-VAC to be secured on top of the mounting rods using a nut. PTFE spacers were used between the housing of the S-VAC and the mounting rods to minimize heat transfer from the oven and heater block assembly attached to the mounting plate. A stainless steel cylinder was machined along with a circular plate with a mounted screw thread. The underside of the cylinder was threaded to match the mounted screw thread on the plate. This allowed the cylinder to be mounted on the base plate. Two holes were drilled through the stainless steel cylinder with the same diameter as that of the annular heat pipe and the cartridge heater. A small recess was drilled for a thermocouple to be placed inside, as part of the heater controller for the heating cartridge.

Drawings of the machined parts that make up the DMS mount and the annular heat pipe can be seen in Figure 3-2 and Figure 3-3.

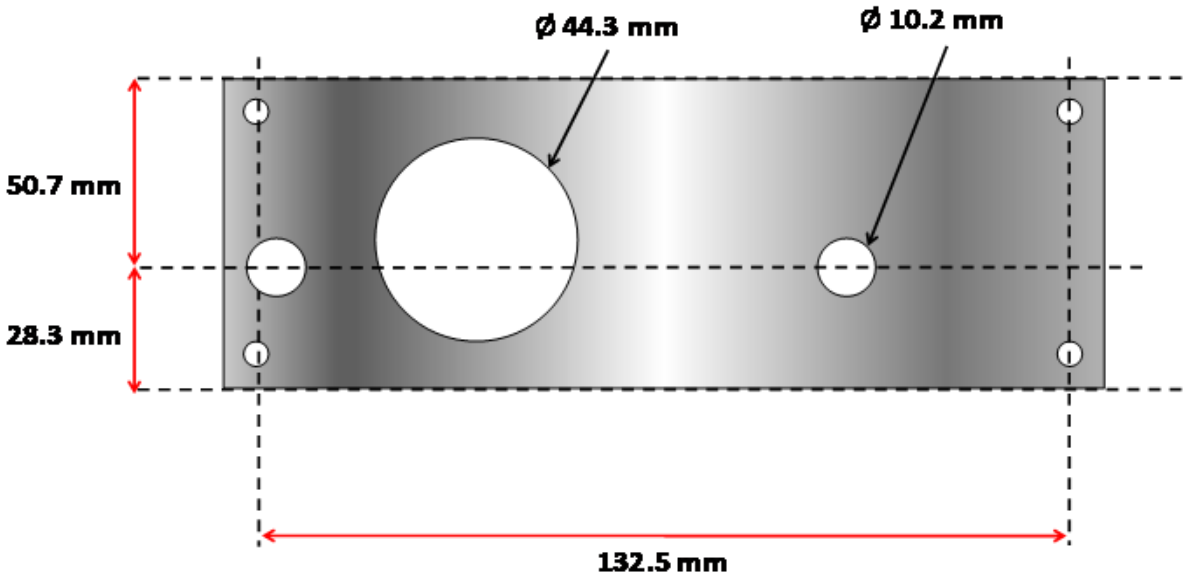


Figure 3-2 Base plate for the differential mobility spectrometer mount. This was designed so that it would fit on the detector mountings of the Varian GC-3800.

Figure 3-2 shows a scale drawing of the base plate of the DMS mounting. The plate can be easily attached to the top of the Varian GC-3800 with no modification of the GC required. All other machined parts are attached to the base plate by screw threads, this means that they can be removed and the base plate can be left attached to the GC oven, should the DMS unit be removed.

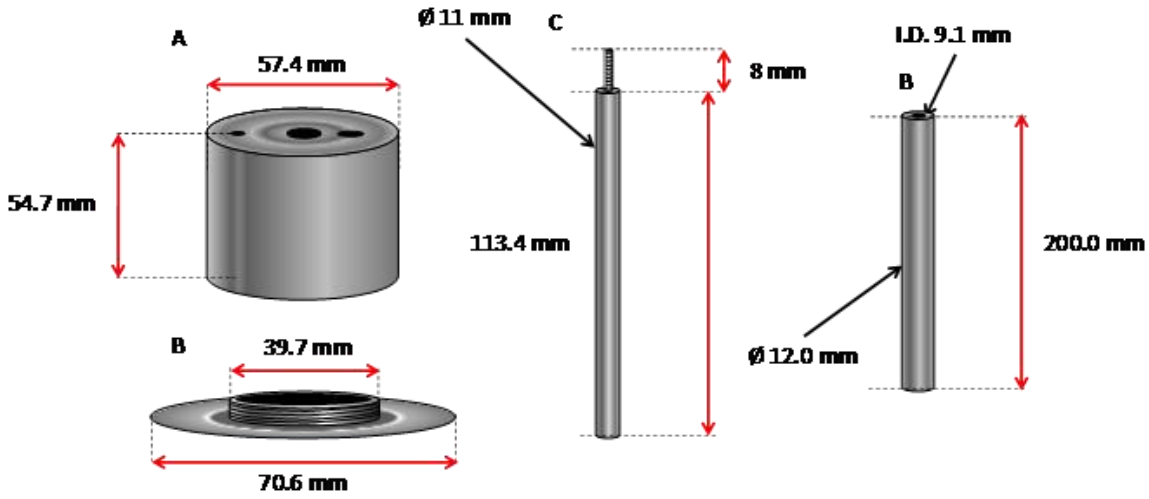


Figure 3-3 The machined parts that fit to the base plate of the differential mobility spectrometer mount. Drawing not to scale.

Figure 3-4 shows the machined parts connected to the base plate. All of the parts were machined from stainless steel. The mounting rod is shown with the threaded rod inserted into the drilled hole, to show the complete rod. This was done to improve the strength of the mounting point.

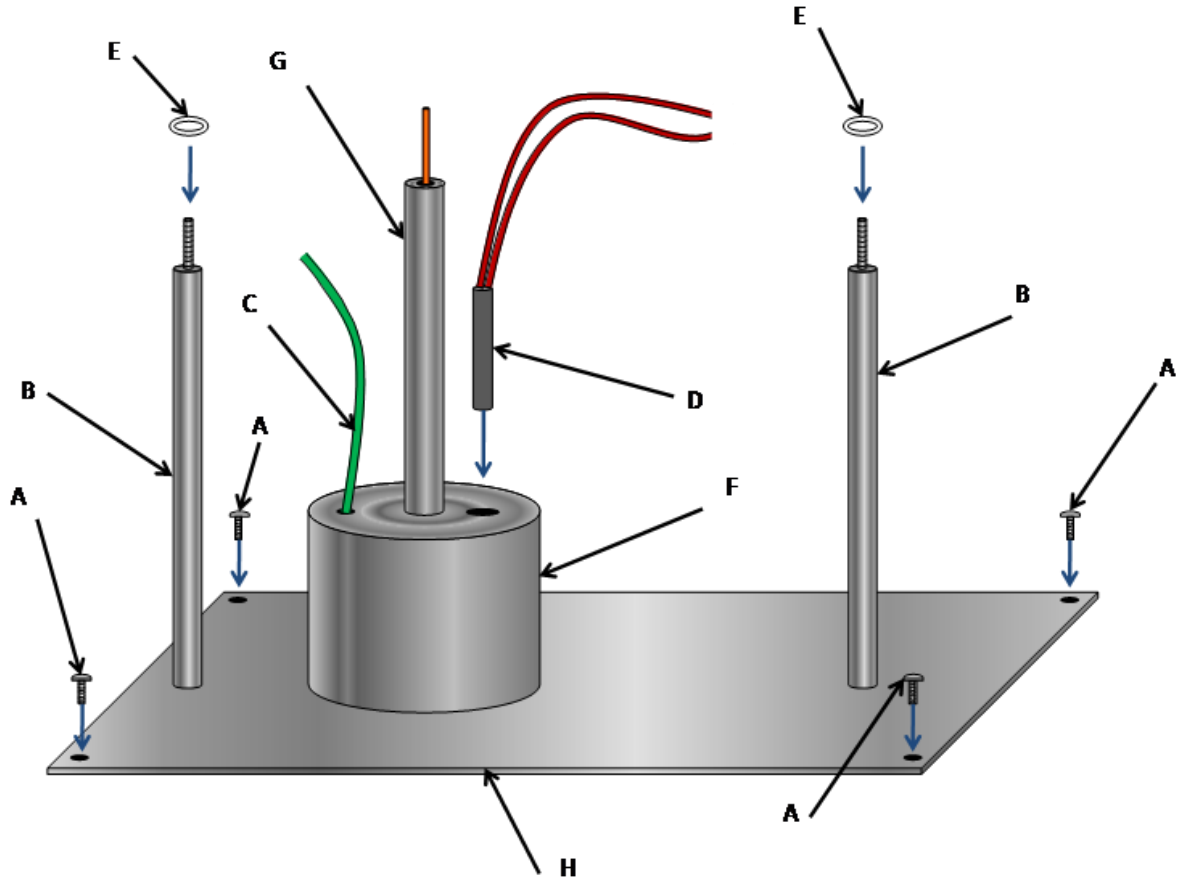


Figure 3-4 Differential mobility spectrometer mount. The blue arrows indicate parts that are fitted together on the mounting of the differential mobility spectrometer. A = anchoring screws, B = differential mobility mounting poles, C = thermocouple, D = cartridge heater, E = PTFE spacer, F = heater block, G = annular heat pipe, H = base plate. The whole mount is retrofit and requires no modification of the Varian 3800-GC to fit.

Once all the parts were machined, the mount was assembled as shown in Figure 3-4 and fitted to the top of the GC. A deactivated fused silica capillary column (0.25 mm I.D (Supelco, USA)), connected to the end of the analytical column, was fed through the annular heat pipe and connected to the Sionex S-VAC DMS. A transfer-line was used rather than the column to maintain the whole length of both columns at the same temperature. This would not have been possible if the column was connected directly to the DMS unit as the annular heat pipe

was not capable of reaching the same temperatures as the 3800-GC oven; this was to protect heat sensitive components within the Sionex system. The mounting points on the S-VAC were then placed over the threaded inserts in the end of the mounting poles and PTFE washers and 3 mm nuts were used to lock the unit in place. The heating block and annular heat pipe were wrapped in Superwool 607 max blanket (RS part No. 417-6713) and this was taped in place on the heating block and held in place on the annular heat pipe using 25.4 mm I.D. black heat-shrinkable tubing wrap (RS part No. 700-4668). The whole block was grounded through the earth point on a mains cable.

Figure 3-5 shows a photograph of the base plate, with the S-VAC unit attached, connected to the top of the GC. The main parts have been labelled with letters corresponding to the drawing in Figure 3-4.

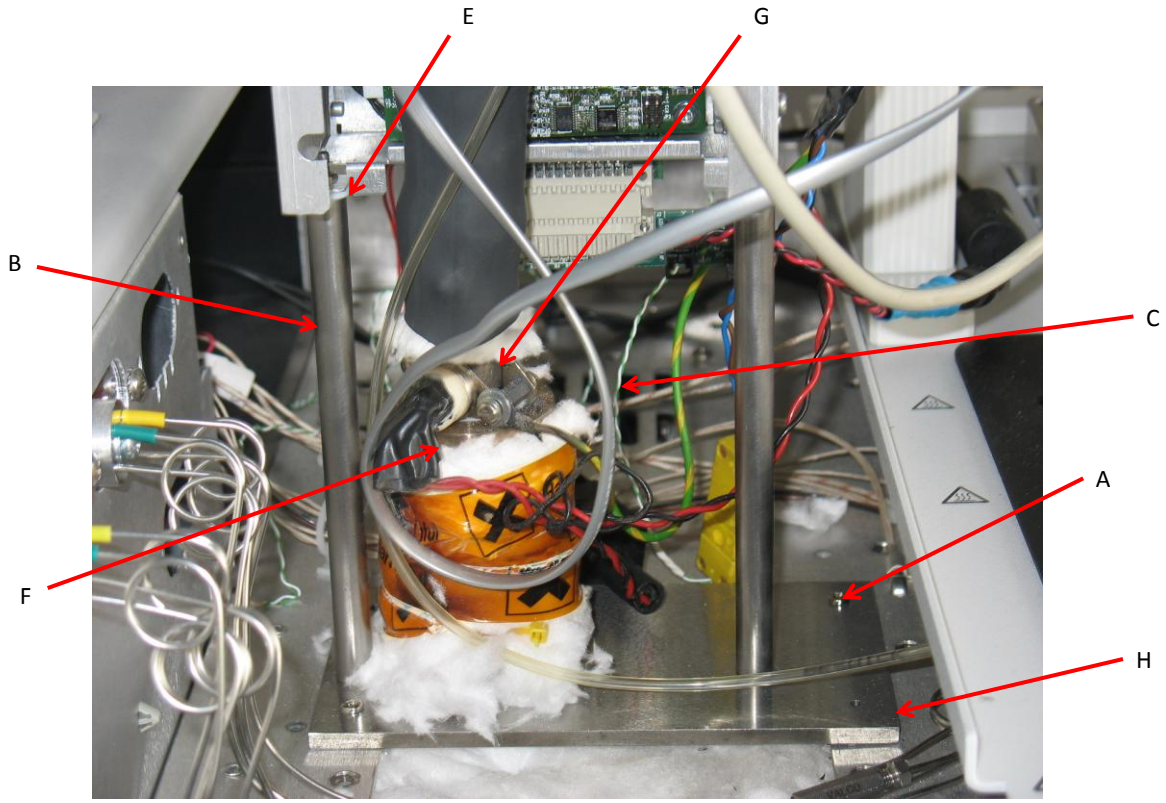


Figure 3-5 Photograph of the differential mobility spectrometer mount assembled on top of the Varian GC 3800 with a temperature controller and a Sionex S-VAC attached. Labelling kept the same as Figure 3-4 for ease of identification.

The parts list for the DMS mount and the heating assembly is given below:

- **DMS mount**
    - Annular heat pipe - (Pipcar Ltd, Unit 4B Valley Industries, Hadlow Road, Tonbridge, Kent, TN11 0AH. England. Tel (UK) +44 (0)1732 851807. Fax (UK) +44 (0)1732 850255).
    - Baseplate and mounting poles - (Produced externally)
    - PTFE spacers - (Produced by the University of Manchester Institute of Science and Technology workshop)
    - Heat pipe mount with threaded base - (Produced by the University of Manchester Institute of Science and Technology workshop)
    - 3mm thread mounting screws - (internal stock)
- **Heating assembly**
    - Tempatron On/off temperature controller, 0-200 deg C type K. RS stock No. 344-546
    - PTFE insulated J welded tip thermocouple. RS Stock No. 621-2142
    - Bi-metallic NC thermostat. RS stock No. 339-724A
    - 300W Cartridge heater 6.5 x 100mm. RS stock No. 376-1685

Figure 3-6 shows the DMS mount in place on the GC oven.

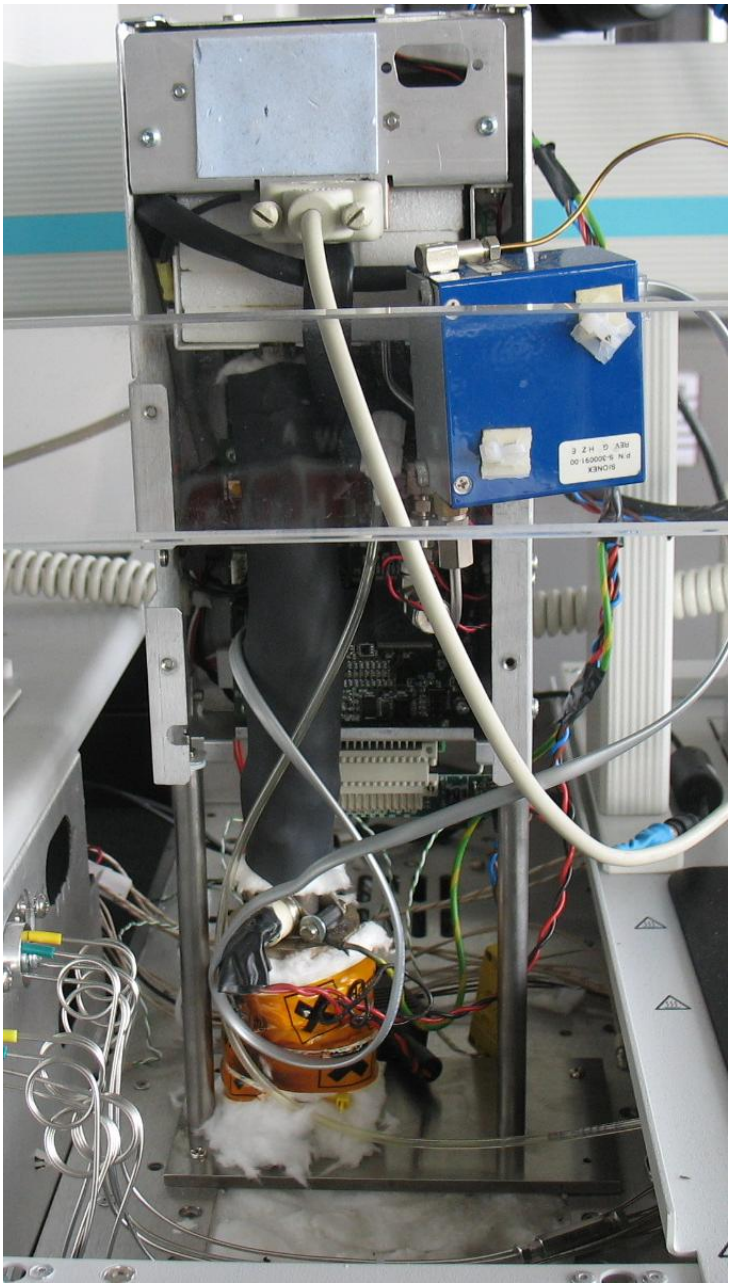


Figure 3-6 The DMS mount set up with the DMS in position.

3.2.5 Splitting the sample

A fixed ratio inlet splitter was used to split the thermal desorption profile between the two capillary columns. The thermal desorber transfer line was connected to a VSIS-2 inlet splitter, Figure 3-7, (SGE part No. 123632) with GVF/004 (SGE P/N 072663) ferrule for connection of the thermal desorber fused silica capillary and a GVF2/004 (SGE P/N 072662) ferrule for connection of two Agilent J&W DB-5ms (I.D. 0.25 mm, 30 m, 0.5 µm Agilent part No. 122-5536) analytical columns inside the gas chromatography oven.

Figure 3-7 Column splitter

3.2.5.1 Mitigation of differences in transfer line temperatures.

The heated transfer line to the mass spectrometer was specified at higher and more precisely maintained temperatures than the annular heat pipe. To avoid temperature effects on retention times between the two channels deactivated, fused silica capillary column was used to connect the DMS to the main capillary column, using a SilTite™ mini-union GC capillary column connector (SGE Part No. 073550) with two 85 per cent Vespel/15 per cent graphite 0.1-0.25 mm I.D. ferrules (SGE Part No. 072696). This fused silica column was passed through the annular heat pipe with the SilTite™ mini-union maintained inside the GC oven. This fused silica column contained no stationary phase so analytes would pass along it without any partition effects. This meant that while there would be a difference in temperature between the two, the analytes would move quicker through the deactivated column compared to the main capillary columns and so reduce any differences in retention times. The same set-up was used for the mass spectrometer.

3.2.5.2 Balancing flows.

Engineering a precise 1:1 split between the two columns was not straightforward. The priority focus was to deliver a flow of 2 mL min^{-1} through the mass spectrometry column for two reasons;

1. Low pressure at the end of the capillary column damages the stationary phase and causes contamination of the mass spectrometer, and a minimum of 1 cm^3 min^{-1} should be maintained and,
2. As the mass spectrometer provides identification of the analytes and this can be confirmed with retention time against a standard, every effort should be made to keep this as consistent as possible when prospecting for biomarkers.

In order to balance the flow in this way the deactivated capillary retention gaps to the two detectors were trimmed to the lengths needed, and this in turn meant that the effects of temperature on column flow were not identical in each of the channels. This meant that the flow through the DMS column was unlikely to exactly match the flow in the MS column, it was still raised at the same rate during each analysis and the results between the detectors were directly comparable.

3.2.5.3 Calibrating the flow through the columns

To achieve the specified flow of 2 cm^3min^{-1} over the specified temperature range, measurements were made at 10°C steps from 30 to 310°C at varying pressures to produce a series of charts; an example is shown in Figure 3-8.

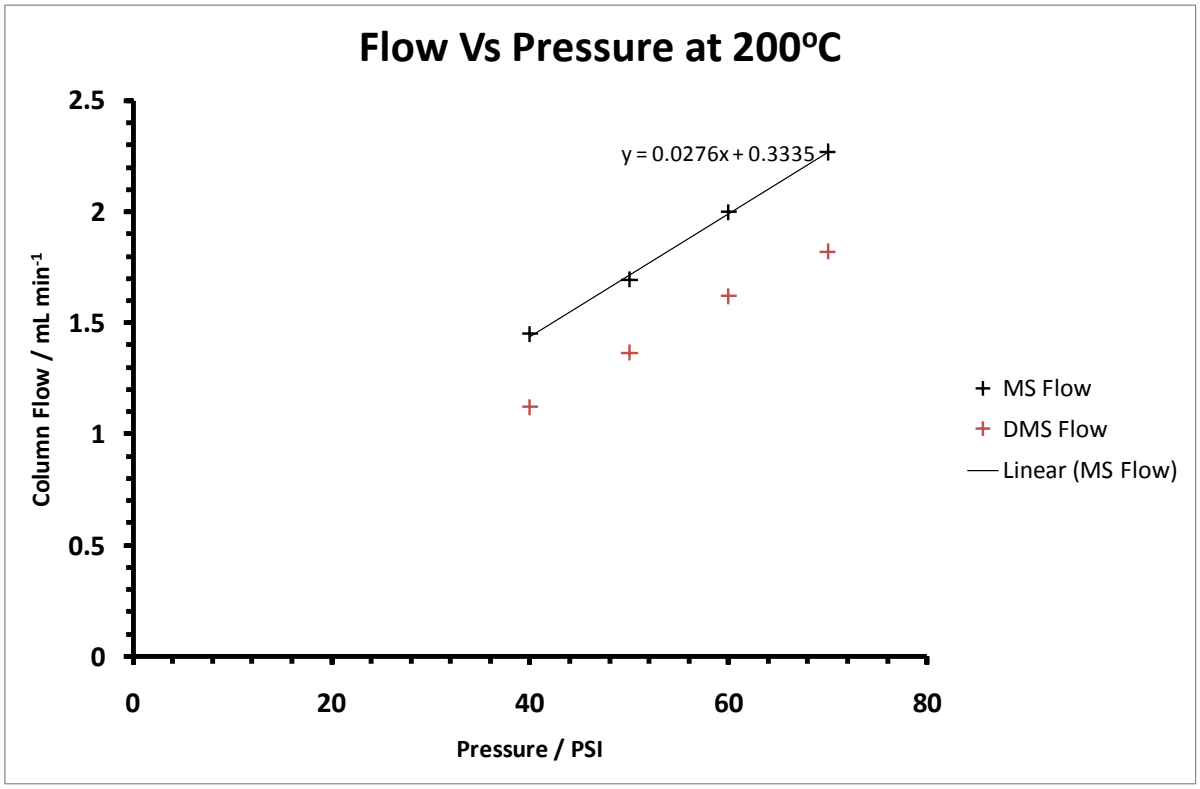


Figure 3-8 An example of the charts produced to show how flow varied with pressure at a constant temperature.

Figure 3-8 shows the change in flow to be linear with respect to pressure at a constant temperature. To show that the change in flow was also linear with respect to temperature at constant pressure, another series of charts were plotted, an example of which is shown in Figure 3-9.

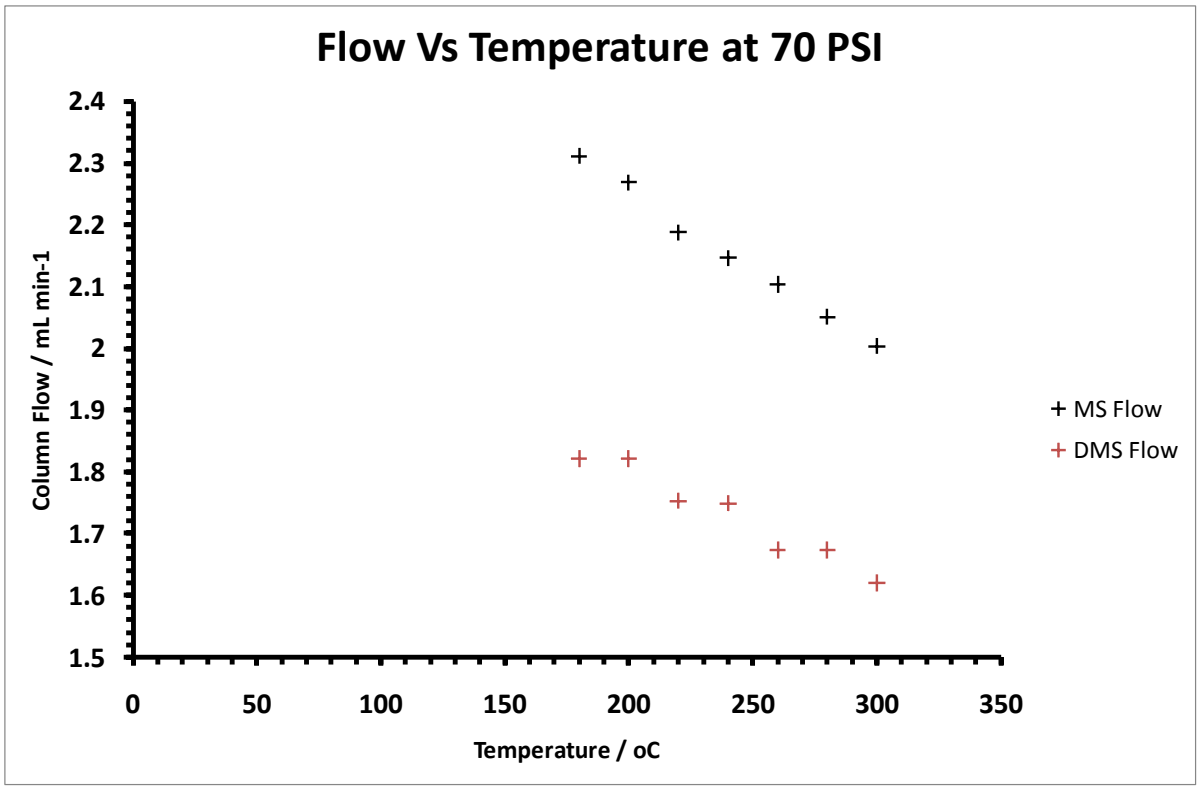


Figure 3-9 An example of the charts produced to show how flow varied with temperature at a constant pressure.

Figure 3-8 and Figure 3-9 show that the variation of flow is linear with respect to both temperature and pressure. This means that the equations of either line can be used to calculate the temperature and the pressure required to maintain the flow at a constant 2 cm^3 min^{-1} for any time point in a GC programmed run. Figure 3-10 shows a plot of pressure against temperature to give a flow of 2 cm^3 min^{-1} at each temperature point that measurements were made.

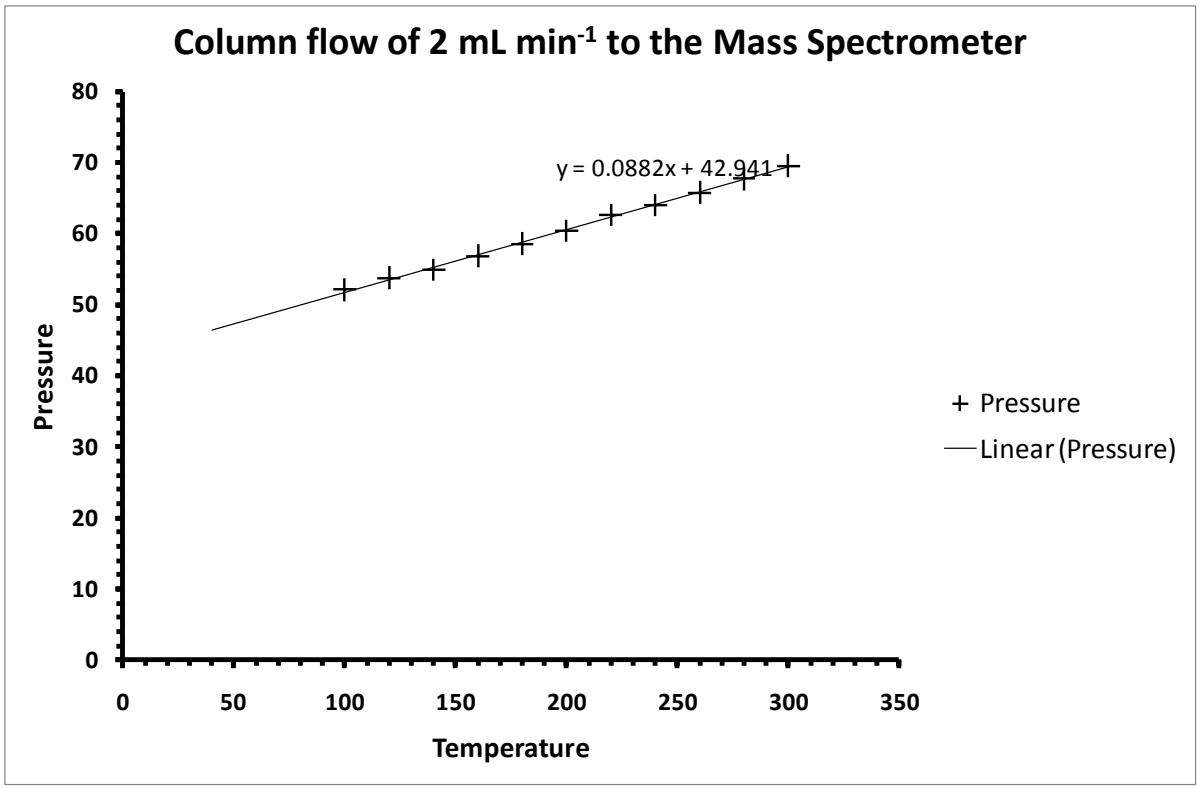


Figure 3-10 Chart showing the pressure required to maintain a 2 mL min^{-1} to the mass spectrometer.

The equation of the line in Figure 3-10 allows for the calculation of the pressure required for a flow of 2 cm^3 min^{-1} at any temperature within the range measured.

As can be seen in Figure 3-8 and Figure 3-9, the flow through the DMS column is lower than that through the MS column. This means that the same analytes will reach the two different detectors at different times, highlighting one of the problems with a parallel column approach. This must be taken into account and corrected for. This can be done using a retention index ladder. While this difference exists it will be constant across all of the samples analysed and so they will still be comparable.

This procedure needs to be carried out each time a column is replaced as slight differences in the two columns may change the pressure required to drive the correct flow through them.

3.2.6 System operating parameters

The system operating parameters are given in Table 3-1, Table 3-2, Table 3-3 and Table 3-4. These are the parameters used for the analysis of all breath samples in this work.

Table 3-1 Thermal desorber settings

Parameters	Settings	
Primary desorption flow	50	cm^3min^{-1}
Primary split	splitless	
Primary desorption temperature	300	^oC
Primary desorption time	5	minutes
Cold trap temperature	-10	^oC
Secondary desorption flow	20	cm^3min^{-1}
Secondary split	Splitless	
Secondary desorption temperature	300	^oC
Secondary desorption time	5	minutes
Trap heating rate	100	$^oC\ min^{-1}$
Transfer line temperature	140	^oC

Table 3-2 Gas chromatography settings

Parameter	Settings		
Column	95% methyl 5% phenyl 30m x 0.25mm I.D x 0.25um film thickness		
Carrier gas	Helium		
Flow	2 cm^3 min^{-1}		
Total run time	60 minutes		
Temperature program	Temperature/ ^oC	Rate/ ^oC min^{-1}	Hold / minutes
	40	3.3	-
	90	2.5	-
	140	10.0	-
	310	0.0	8.85

Table 3-3 Mass spectrometer settings

Parameter	Setting
Scan type	Full
Mass range	m/z 40-450
Ionization time	Electron ionization (EI)
Emission current	10 μA
Scan time	0.82 s
Transfer line temperature	270^{o}C
Trap temperature	150^{o}C
Manifold temperature	50ºC

Table 3-4 Differential mobility spectrometer settings

Parameter	Settings
buffer gas	High purity nitrogen
Flow buffer gas	270 cm^3min^{-1}
Temperature	100ºC
Compensation field scan rate	3.6 s $scan^{-1}$
Dispersion field scan rate (dispersion field programming)	322 V $cm^{-1}s^{-1}$

4 Processing GC-DMS data surfaces. A new Euclidian-based approach

4.1 Introduction

The data generated by GC-DMS are difficult to process in multivariate analysis, and this challenge was a vital part of this study if DMS devices are to be used in this field. DMS data is a plot of compensation voltage versus intensity, but when a DMS instrument is attached to a GC system an extra axis is introduced creating a data surface, and including data from multiple samples results in a data cube. Multivariate techniques, such as principal component analysis or principal component – discriminant function analysis are the favoured approach in such instances; however the difficulty with working with GC-DMS data is that multivariate techniques can only be applied to one dimensional data. It is possible to reduce DMS data to one dimension by integrating across the compensation field, but this removes any separation information generated by the instrument in the first place. For complex data sets such as breath samples this lost information may be vitally important.

This chapter therefore describes a method that was developed to register the three-dimensional output generated by a GC-DMS experiment as a two-dimension array amenable for multivariate processing.

4.2 DMS data processing

IMS has been developed for applications where analytes are specified and their behaviour within the instrument is characterised extensively [69-76]. Subsequently the instrument's parameters are optimised to detect the specified compounds. In contrast in this study DMS was used to detect unknown VOCs in a complex mixture, and the instrument was configured to record as much information as possible with the greatest possible sensitivity. At the end of the campaign the data were to be modelled to enable differences in the sample sets to be evaluated as predictors of disease. This is a "classic" multivariate problem with small, yet significant, differences embedded in highly complex and variable data sets [26, 46, 82, 152, 153].

This study generated many thousands of time resolved spectra, described in three dimensions by:

Compensation field (V cm^{-1});

Signal intensity in (V) and;

Retention time (s) or dispersion field (V cm^{-1}) strength, which correlated to retention time as the dispersion field was being programmed [109].

The first step in processing such data is to reduce the dimensionality of the data structure for example:

Sum the data across the compensation field axis [46, 152], this removes the separation given in the compensation field and can combine peaks that were resolved,

or, take strips through the data at specific compensation voltages [153]. For complex data sets this either means reducing the amount of data or the complementary processing of a very large number of strips,

Or, to use wavelet analysis to pick out peaks of interest and process these [154]. This can also reduce dimensionality of the data as peaks below the limits set in the analysis may be missed.

Summed method approaches integrate across one of the dimensions of either retention time or compensation field, and in recent approaches the preferred factor has been compensation field [46, 152, 153], a schematic of how this is done is shown in Figure 4-1. This approach removes chemical information contained within the differential mobility dimension and necessarily excludes the reactant ion peak. This is a significant omission, for earlier studies with neural network analysis of mobility spectra indicated that significant information is contained in the fragment ions associated with the RIP [155]. Therefore intentionally excluding this element of the experimental response is to remove a significant source of information from the data analysis.

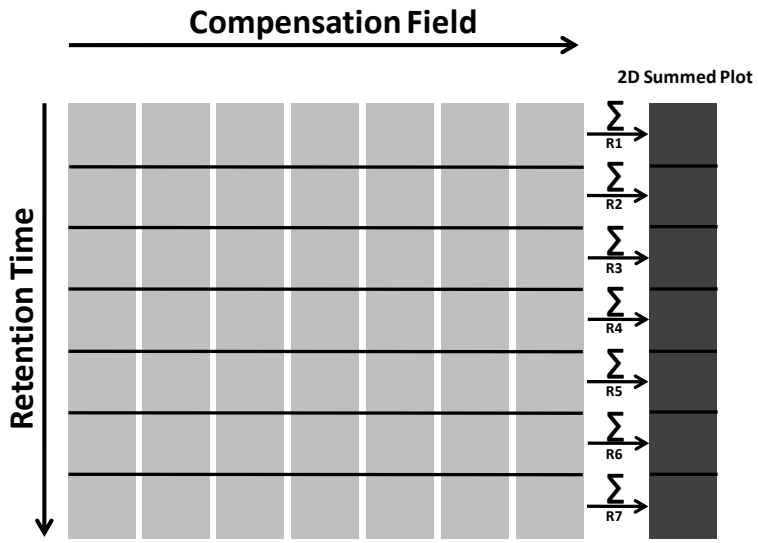


Figure 4-1 Representation of how the summed method transforms the data from three to two dimensions. The raw data is summed across each compensation field sweep and the resultant data is used.

Ideally the data structure needs to be simplified while retaining as much of the information as possible. One approach may be to employ a Euclidian approach; whereby multiple dimensions may be replaced with single vector.

4.3 Developing the approach

4.3.1 Euclidian distance method

The Euclidian distance method represents the vectors of compensation field and retention time by a single vector. For a three-dimensional response surface of N values of compensation field and H values of retention time then any point on the surface may be represented by the single vector Z_x (Where x is the specific point on the surface) and I, which is the intensity of the signal at the specific point. This vector is calculated according to Equation 4-1, and a schematic detailing how a point is identified is shown in Figure 4-2.

$$Z_x = \sqrt{{n_i}^2 + {h_j}^2} \qquad \textbf{Equation 4-1}$$

Here, n_i is the scaled value of the compensation field between 0 and n and h_j is the corresponding scaled retention time increment between 0 and h.

To enable this approach to work, two-additional data processing operations are required to take into account the original "scaling" of the data set and "equivalent-value" Euclidian vectors that result.

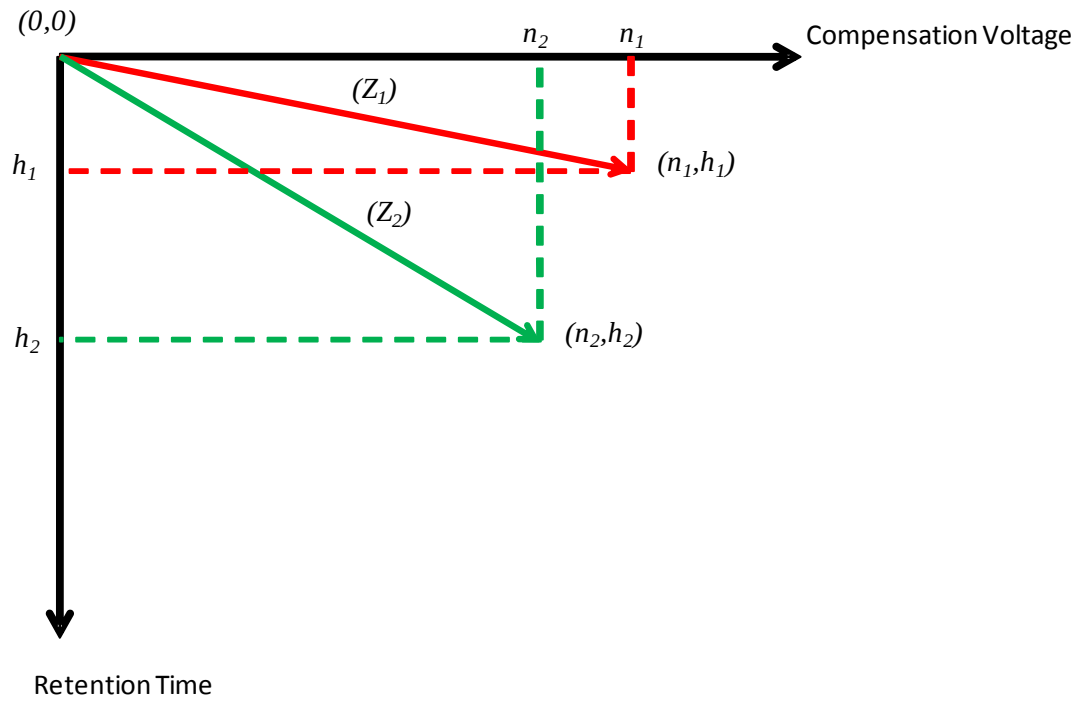


Figure 4-2 Schematic showing how each point is identified on a surface for the Euclidian transformation.

4.3.2 Data scaling

The "raw" data are defined by the experimental parameters (compensation voltage and retention time). Combining data scales with contrasting numerical ranges generates a matrix that is disproportionately skewed towards one of the variables, with the associated reduction in information. Further, the inclusion of un-scaled positive and negative compensation field values would result in nonsensical outcomes. Consequently the data are scaled from zero to $\boldsymbol{N}$ and $\boldsymbol{H}$ in single integer series, where zero is the lowest value of the measurement scale number and $\boldsymbol{N}$ and $\boldsymbol{H}$ are the highest values of the measurement scale, $\boldsymbol{N}$ being associated to the compensation voltage scale and $\boldsymbol{H}$ being associated to the retention time scale.

4.3.3 "Equivalent value" Euclidian vectors

The mapping of a surface using Euclidian distance will produce a series of equivalent-value vectors. The number of equivalent value Euclidian vectors points will depend on the data set and the scaling increments used. This is addressed by adding a point number when sorting the Euclidian/Intensity pairs. As the Euclidian distance for each point is calculated, a point number is assigned to each distance. The distances are calculated across the compensation voltage axis beginning at the origin and proceeding across the axis until every point in this compensation voltage sweep has been assigned. This is then repeated along the retention time axis. By sorting by Euclidian distance first and then point number, features that appear first in the data set will appear first in the reassigned set. The sorted pairs are then reassigned against a single integer series, ranging from 1 to $\boldsymbol{M}$ **(**$\boldsymbol{M=N\times H}$**)** when one is the smallest distance and $\boldsymbol{M}$ is the largest distance. This enables the full information of the set to be retained.

4.3.4 Five peak demonstration test set

An idealised data surface consisting of five peaks was created using a two dimensional Gaussian function;

$$f(x,y) = Ae^{-\left(\frac{(x-x_0)^2}{2\sigma^2 x}+\frac{(y-y_0)^2}{2\sigma^2 y}\right)} \quad \boldsymbol{f(x,y) = Ae^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2}\right)}} \qquad \textbf{Equation 4-2}$$

Where A describes the amplitude of the peak, x_0 and y_0 describe the centre of the peak and σ_x and σ_y describe the x and y spread of the peak. These will be referred to as the peak parameters. As an example of how these factors affect the shape of the peak several plots have been produced in Figure 4-3 with their peak parameters listed in Table 4-1.

Table 4-1 The peak parameters for the different peaks seen in Figure 4-3.

Series number	Intensity (A)/AU	Peak location (x_0 or y_0)/AU	Peak spread (σ_x or σ_y)/AU
1	1	0	3
2	2	0	2
3	4	-4	2
4	5	0	1

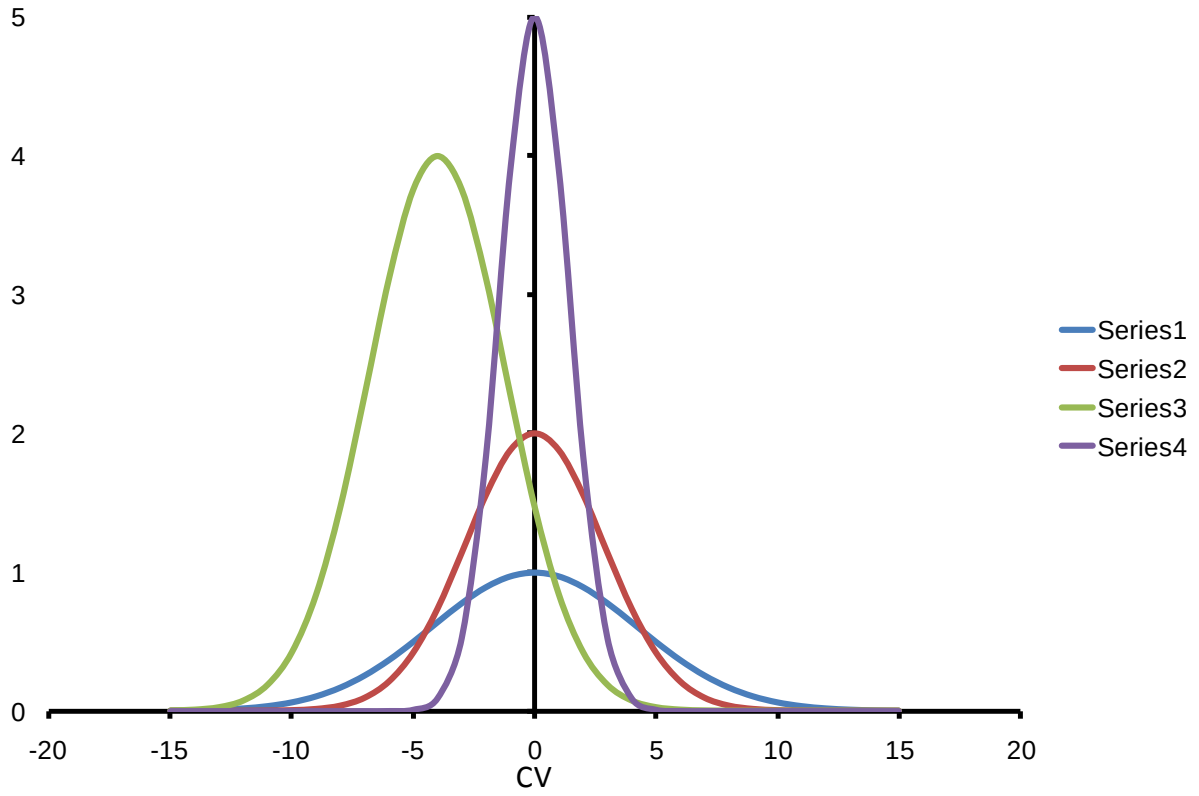


Figure 4-3 Plots of the Gaussian peaks whose parameters are listed in Table 4-1. This shows how altering these parameters affect the size, shape and position of the peak and can be manipulated to created peaks for testing the Euclidian transform method.

Figure 4-3 shows one dimensional Gaussian function plots. In two dimensions the plot has two more variables which dictate the spread and location on the second axis. An example plot is shown in Figure 4-4.

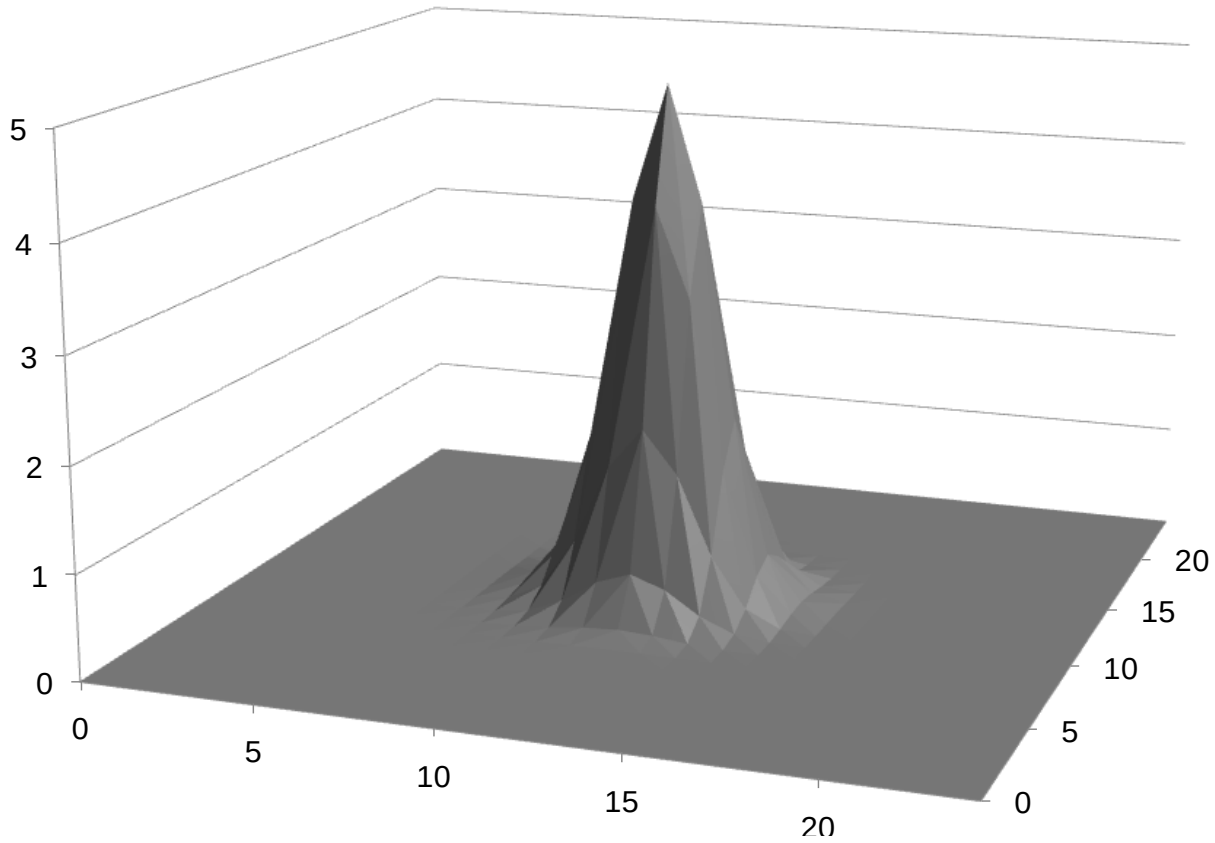


Figure 4-4 An example plot of a single three dimensional Gaussian peak.

The peak seen in Figure 4-4 has the same location and spread along both axes of the plane to produce a symmetrical peak. This can be used to create test DMS data files or insert simulated test peaks into DMS data files.

Five separate peaks were created using Equation 4-2. Their locations were chosen so that two pairs of peaks were unresolved in the retention time axis and analogously, two were unresolved in the compensation field axis. The final peak was placed exactly in the middle of the two pairs. This can be seen in Figure 4-5**.**

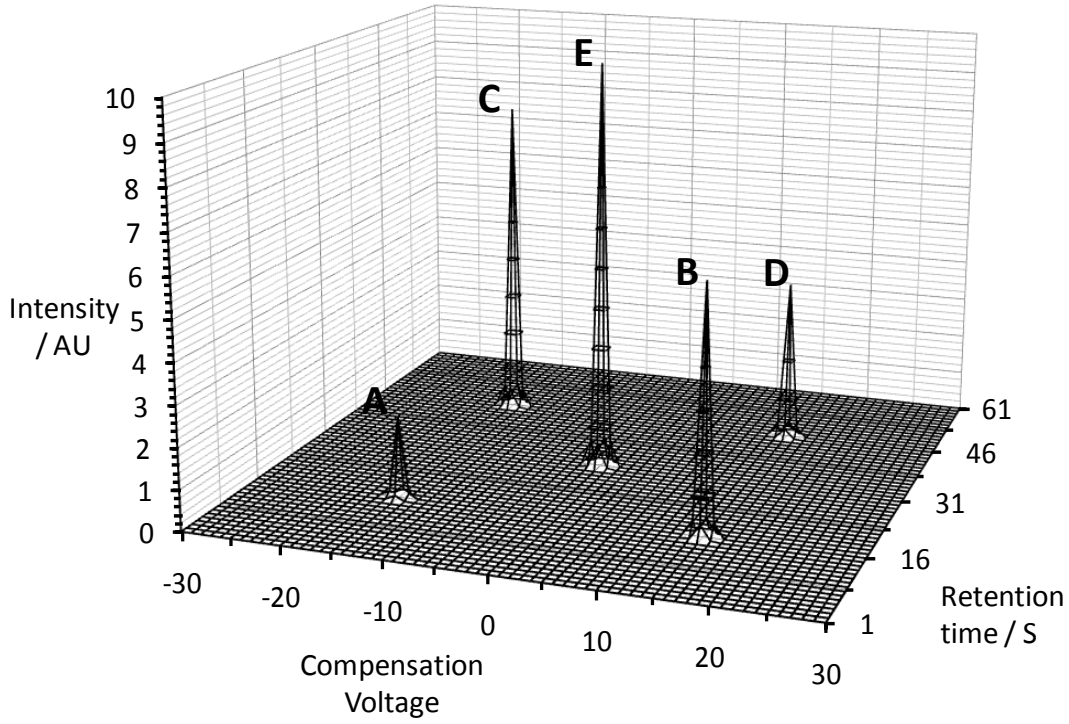


Figure 4-5 Plot of the five peak test set created using the two dimensional Gaussian function.

Although the five peaks are resolved two of them, **B** and **C**, have equivalent-value Euclidian distances and four of them (**A**, **B**, **C** and **D**) are aligned in pairs with respect to compensation voltage (**A**, **C** and **B**, **D**) and retention time (**A**, **B** and **C**, **D**). Integration with respect to either compensation field or retention time will yield three peaks, integration with respect to compensation voltage shown in Figure 4-6, with the associated loss of information.

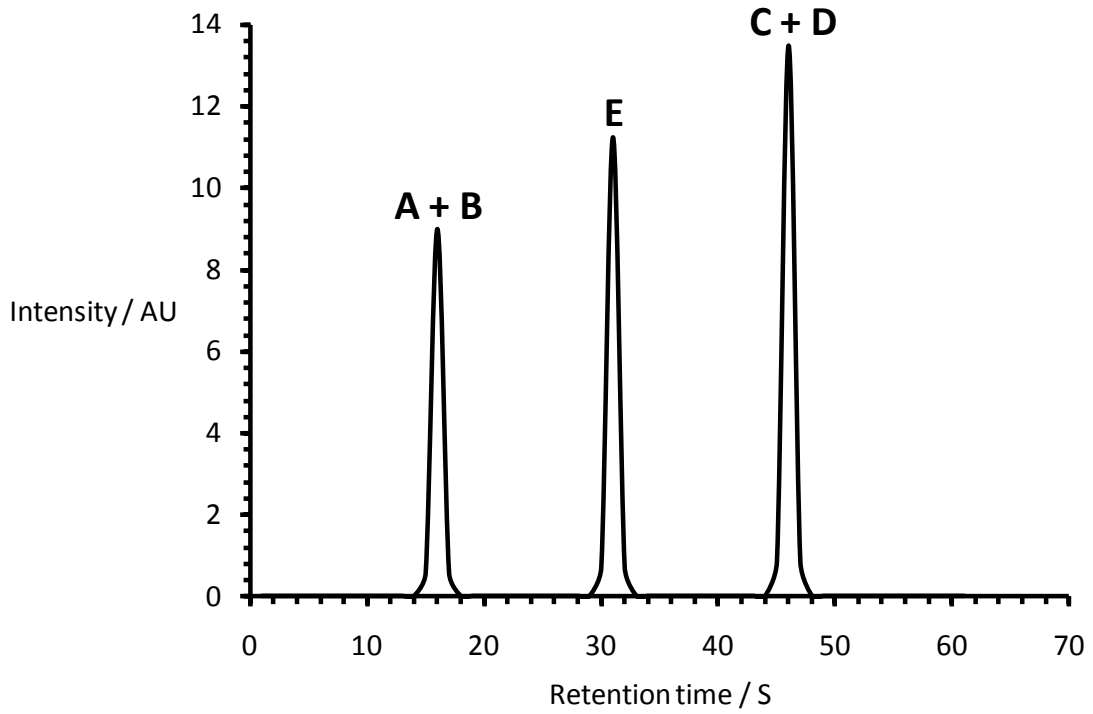


Figure 4-6 Five peak test set after transforming from three dimensions to two dimensions using the summed method with respect to the compensation voltage. Only three peaks are observed with two being enlarged. The peaks are labelled to show which peaks from the original test set make up which peak.

Unfolding using Euclidian distance, data scaling and sorting enables the complete characteristics of all five peaks to be recorded in a two dimensional plot. The smaller peaks, labelled $\mathbf{W_A}$, $\mathbf{W_B}$, $\mathbf{W_C}$, $\mathbf{W_D}$.and $\mathbf{W_E}$ in Figure 4-7 record the peak width and as such enable the full nature of unresolved signals to be included in subsequent data processing operations. This enables more of the complexity of the original data to be captured in the transformed matrix. Note that their position along the "reassigned distance" axis is due to their distance from the origin in the original data set, and co-located features may not normally be assumed to be close to one another in the original response surface, see features **B** and **C** in Figure 4-7 for example.

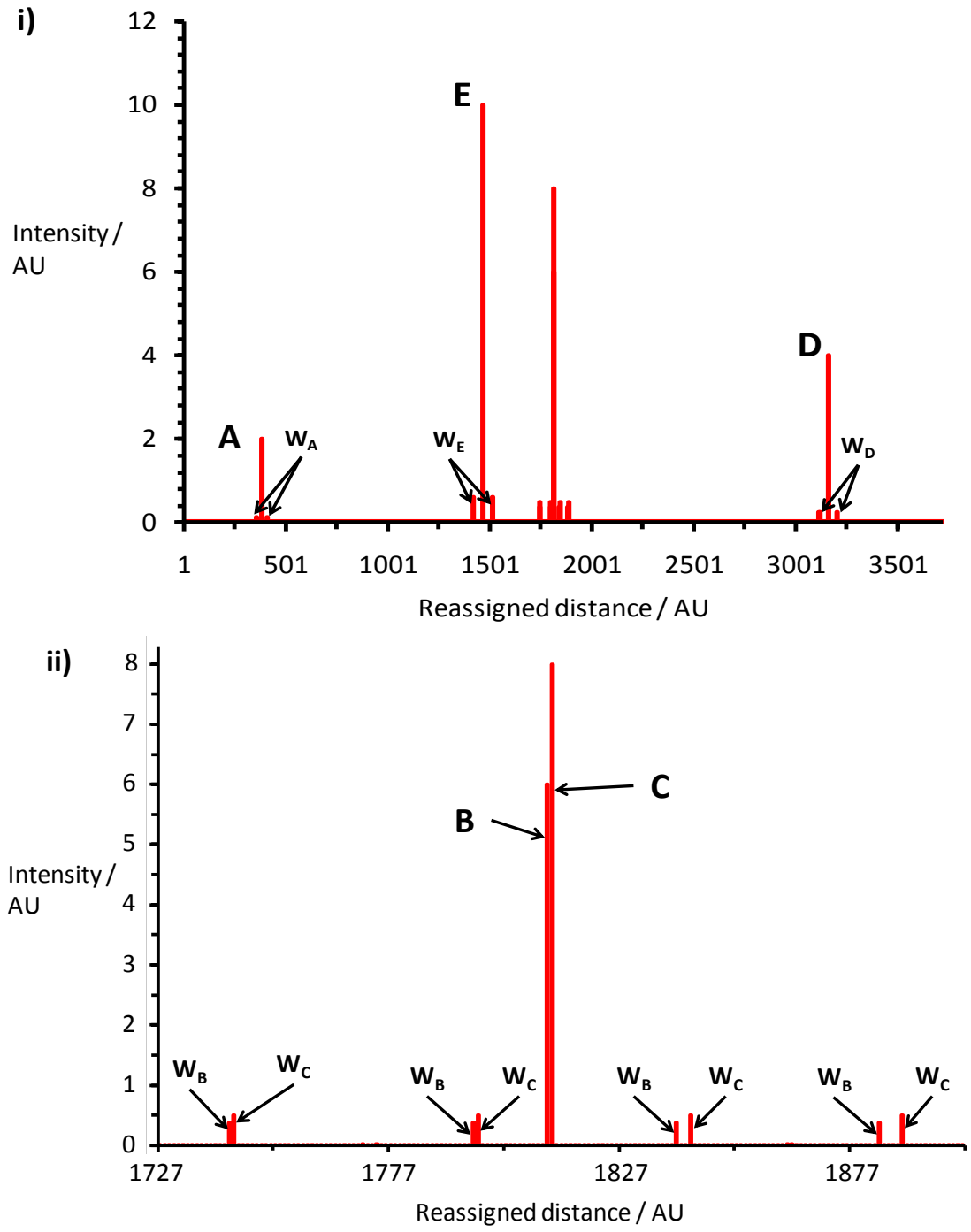


Figure 4-7 Five peak test set after transforming from three dimensions to two dimensions using the Euclidian transform method. Each separate peak is due to a single intensity point in the original data. i) shows the full transformed data set, ii) shows an enlarged view of the reassigned distance region 1727 to 1902. In this region it can be seen that peaks B and C are fully resolved. Peaks labelled W_x are the coded width of the peaks, where x is the peak label.

4.3.4.1 Evaluation of the five peak test set

The five peak test set sought to demonstrate the Euclidian transform approach to data reduction of GC-DMS data. The test set was constructed with overlapping features that would be obscured by a summing/integration approach or by taking sub-samples. This effect can be seen in Figure 4-6 with the first and third peaks both containing pairs peaks which co-elute in the time axis. Any attempt to use summing approaches would lose these details in the subsequent data models. An alternative approach, wavelet analysis, "picks peaks" and sets limits based on pre-determined parameters. While the operator may be very careful this method still reduces the data sets and may exclude important information. Methods where strips are taken through the data have also been used but this means complimentary processing of many strips depending on the complexity of the data set and may not take into account multiple changes in different strips. The Euclidian distance method was applied to the five peak test set; this produced the data set plotted in Figure 4-7. All five peaks can be seen and they are all fully resolved. Smaller peaks are also observed and these are due to the width of each peak. As all the data points in the plot are single signal intensities from the original data matrix there has been no false combination of any components as opposed to the data resulting from using the summed method.

Preserving the intensity of the original signal is discussed in more in section 4.5.1, but as can be seen from the discussion of the combination of peaks in the same compensation field sweep, it is difficult to assign a peak area to a single peak using this method, especially with real data when background noise will be summed as well.

4.3.5 Dispersion field sweep data set

The dispersion field sweep data set consisted of ten dispersion field sweep blanks recorded on consecutive days. The data was generated on a thermal desorption – gas chromatography – differential mobility spectrometry instrument platform, see chapter 3 for instrument details.

The dispersion field sweep data set was produced using the parameters summarised in Table 4-2. The sample set was produced by running the instrument without any sample injection. Each data set was collected in the morning after the instrument had been

validated. The ten samples were collected on ten consecutive days and were visually checked before processing.

Table 4-2 Differential mobility spectrometer settings

Parameter	Setting value	Units
Start compensation voltage	-43	V
End Compensation Voltage	10	V
Number of Compensation voltage steps	100	
Compensation Voltage Scan duration	3600	ms
Start dispersion field voltage	500	V
End dispersion field voltage	1500	V
Number of dispersion field steps	100	
Sensor temperature	80	°C
Flow rate	300	$cm^3 min^{-1}$

4.3.5.1 *Comparison of the effect of using the Euclidian-based approach against a summation method with the summed method to analyse dispersion field programmed data*

The method was applied to a dispersion field sweep data set to contrast the Euclidian distance method and the summed method with straightforward data, in essence an instrument blank

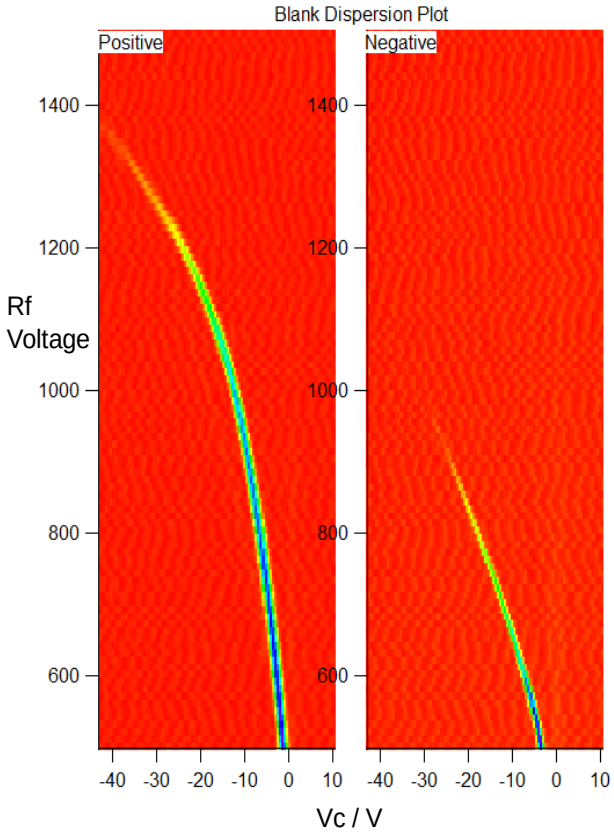


Figure 4-8 Data from the dispersion field sweep data set. The positive channel is displayed on the left and the negative is displayed on the right. The only signal observable in both plots is due to the reactant ion peak.

The dispersion field sweep data set is shown in Figure 4-8. There is only one feature present in both the negative and positive channels and this is due to the reactant ion peak. Note how the intensity of the signal reduces with increasing field strength, due to wall losses, and how the intensity of the RIP in the negative mode is less than the positive mode, transport gasses were dry nitrogen.

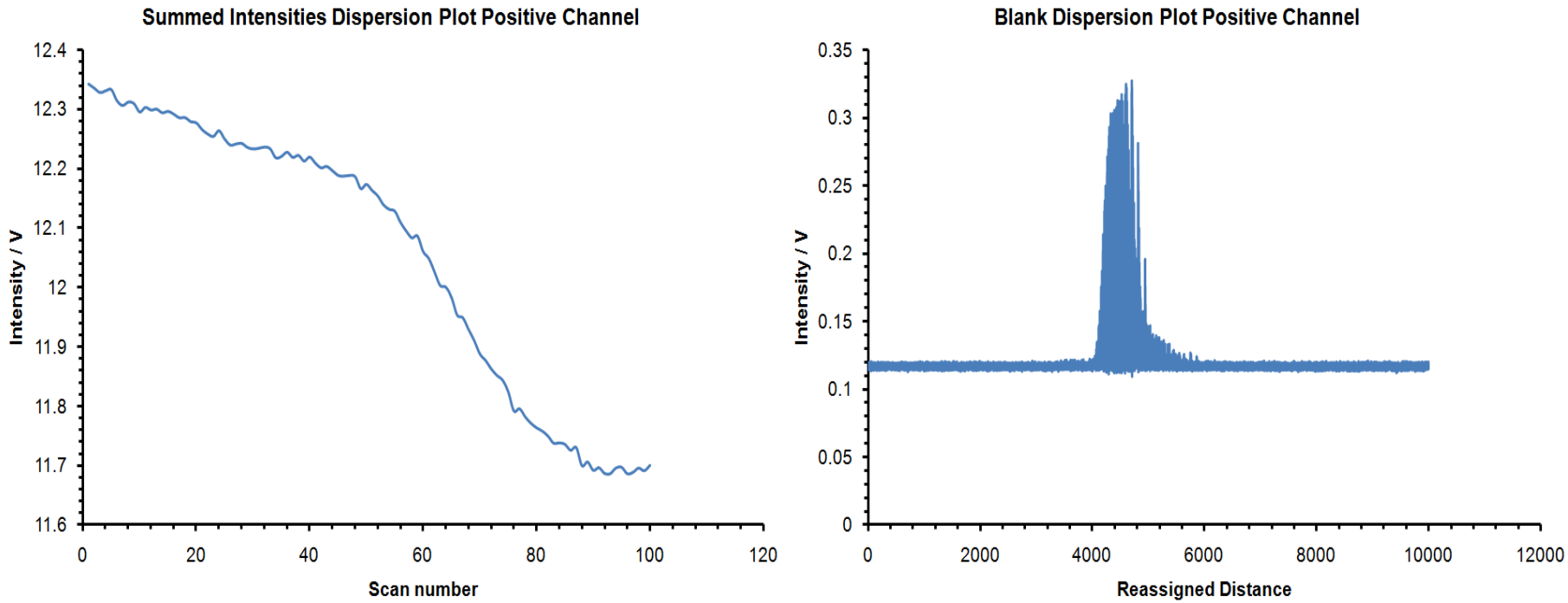


Figure 4-9 Comparison of the final output after converting the positive channel of a single dispersion plot using the summed method and the Euclidian distance method.

The difference in the information content of the two methods is striking, see Figure 4-9. The summed method gives a single line plot that follows the intensity of the reactant ion peak and any underlying water/ammonia features as attenuation of the reactant ion peak with the increasing wall losses associated with increasing dispersion field strength. There are 100 data points in the series, each point being due to a whole, summed, compensation field sweep. The intensity given starts at 12.35 V and falls to 11.7 V. the Euclidian distance method gives a plot where each individual intensity is plotted, and as such is a much larger data set with 10,000 points. The signal due to the reactant ion peak and any underlying peaks are grouped together in the centre of the plot. Figure 4-10 shows an enlarged view of the main response area of the Euclidian distance transform plot.

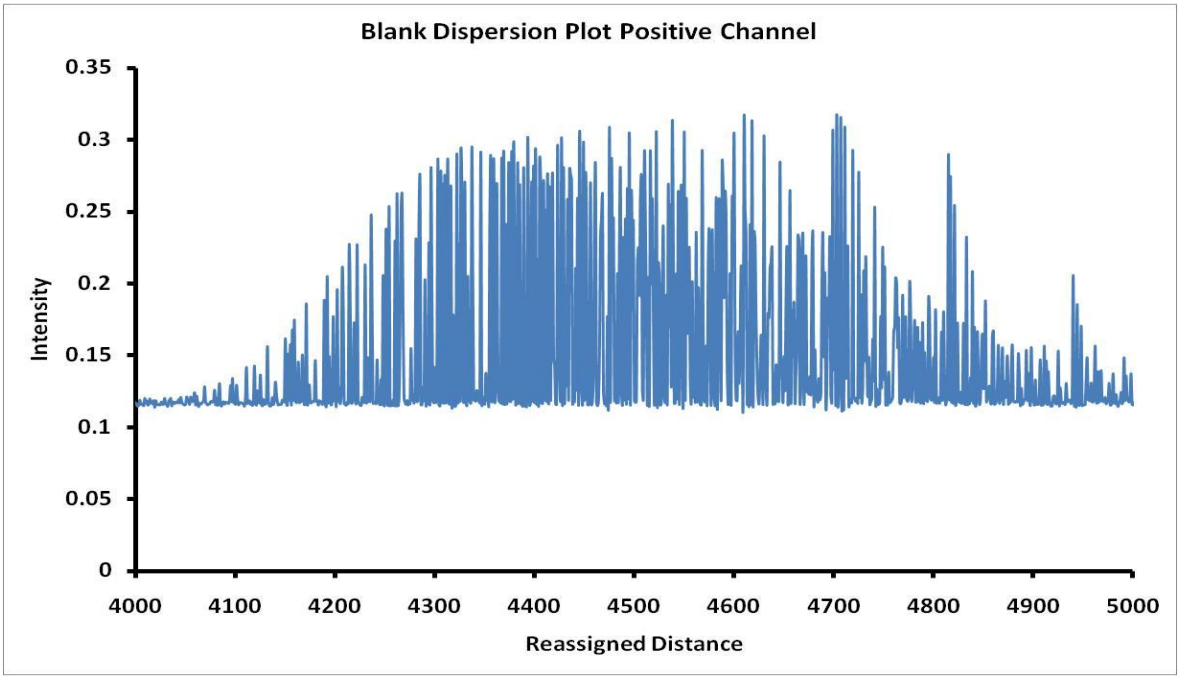


Figure 4-10 Enlarged view of the main response area in the Euclidian distance transform plot of the positive channel.

The decrease in signal intensity in the summed method occurs more quickly in the negative channel data set. The intensity varies from 10.62 to 10.23 V and the data set is the same size as the positive data set. Again, in the Euclidian distance method transformed data we see a grouping of peaks that is more central, although slightly offset to the left, between reassigned distance 2000-4500. The data set is the same size as the positive data set and the signal intensity varies from 0.11 to 0.24 V.

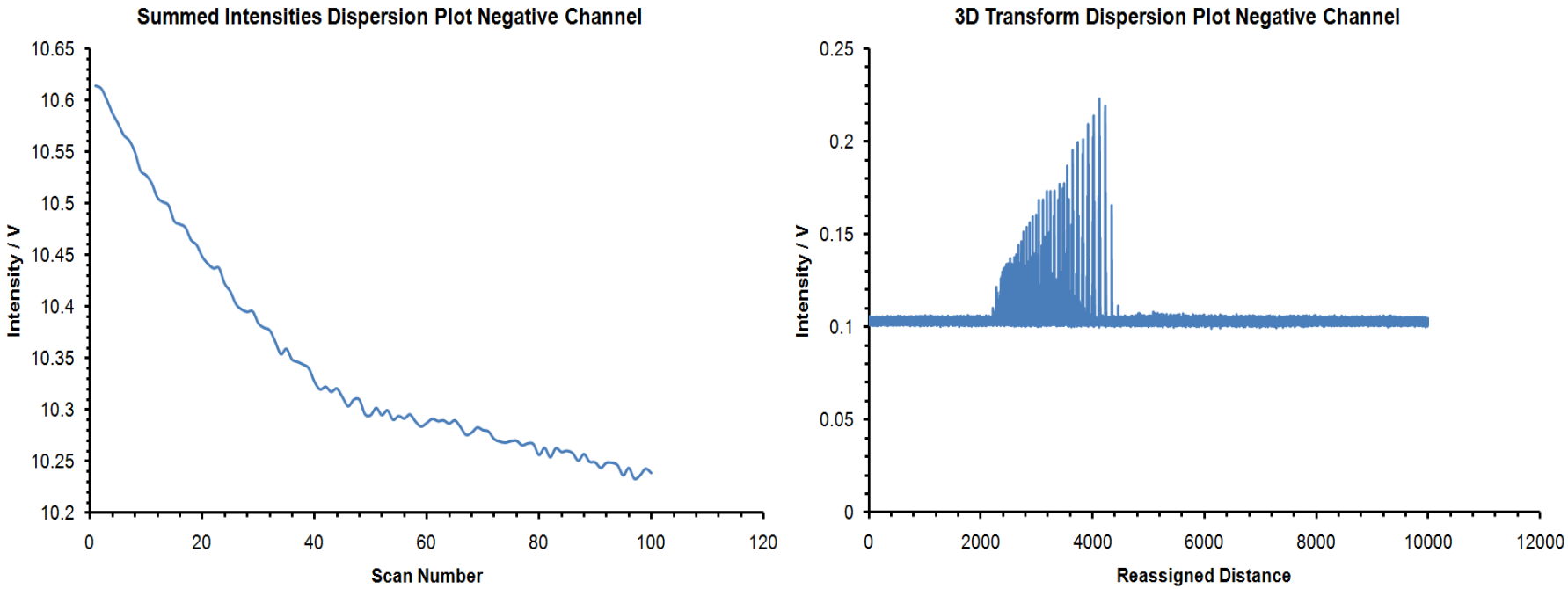


Figure 4-11 Comparison of the final output after converting the negative channel of a single dispersion plot using the summed method and the Euclidian distance method.

An enlarged view of the Euclidian transformed negative channel data is shown in Figure 4-11. The data is less complex than the positive channel data and the transformed features are readily discernable, see Figure 4-12.

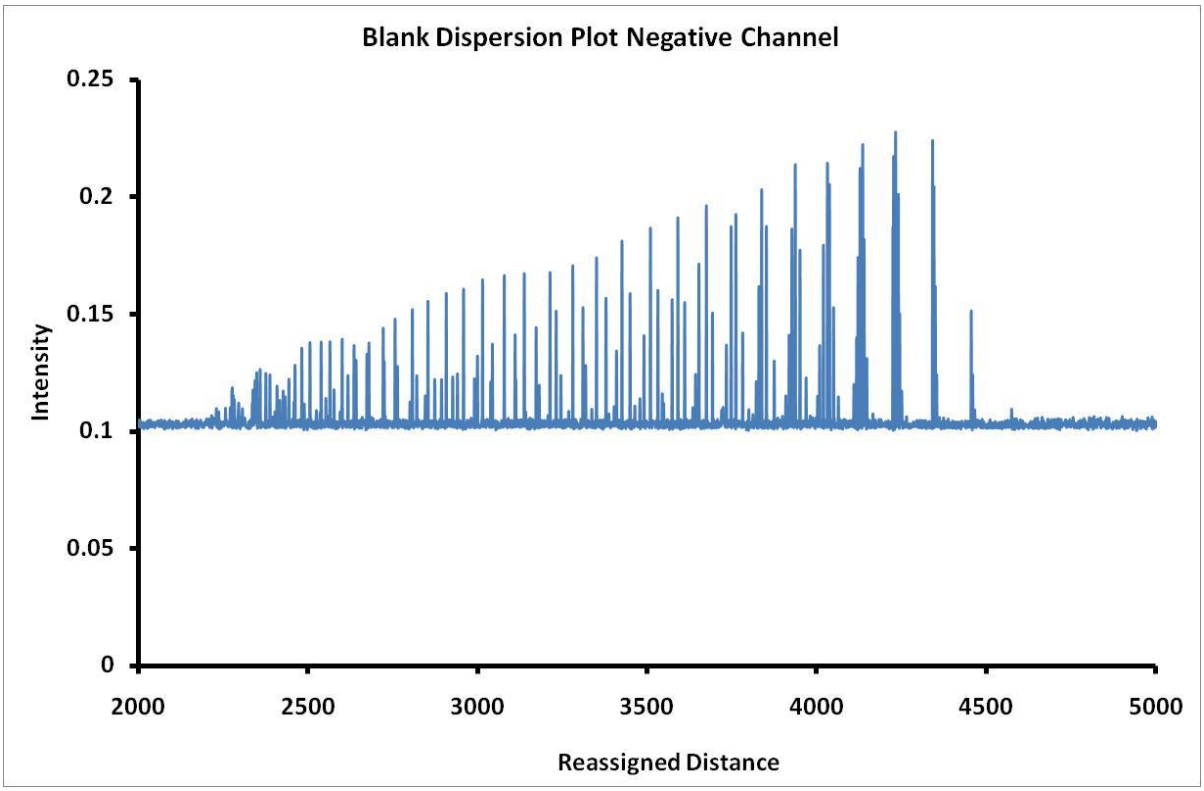


Figure 4-12 Enlarged view of the main response area in the Euclidian distance transform plot of the negative channel.

4.3.5.2 PCA of dispersion field sweep data sets

The ten sets of summed and Euclidian processed dispersion field data were analysed by PCA and the first two principal components were calculated with the data plotted as individual principal component values. The data sets were labelled as either positive or negative. This can be seen in Figure 4-13.

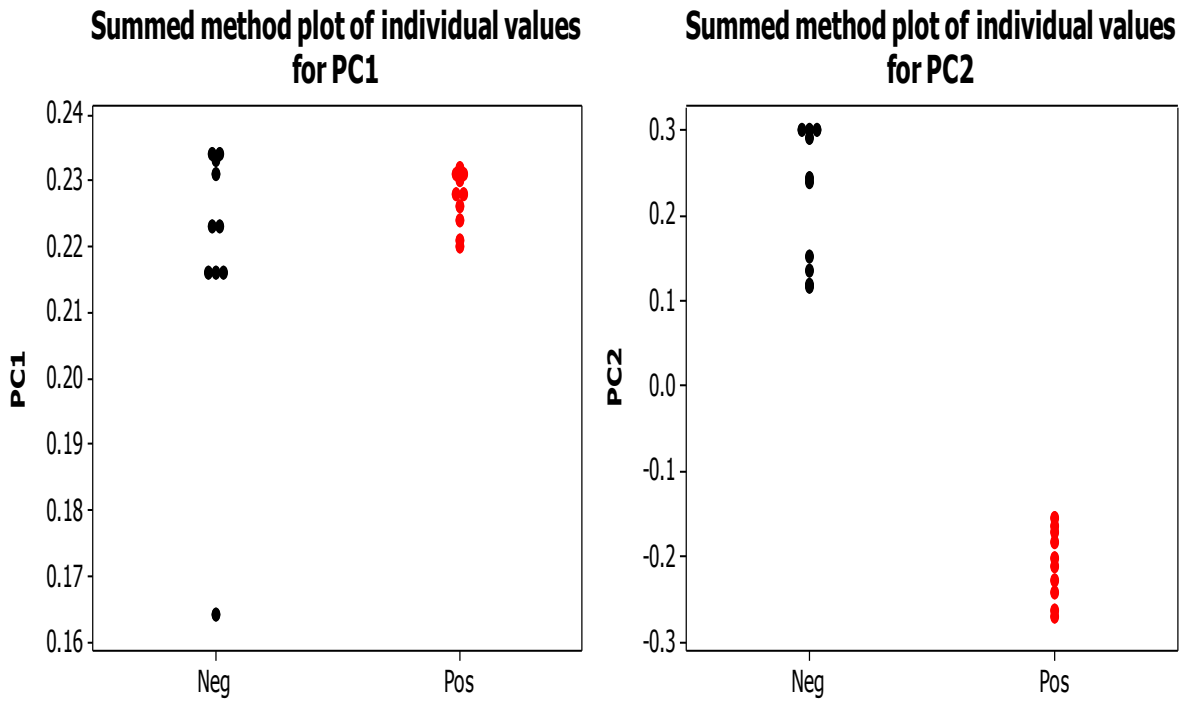


Figure 4-13 Plot of the individual values of the first and second principal components calculated from the data set after transformation using the summed method.

As can be seen in Figure 4-13, when using the summed method for transforming the data, full separation of the data is only observed in the second principal component. Figure 4-14 shows the single principal component plot for the first two principal components calculated from the data after it has been transformed using the Euclidian distance method. Full separation can be seen in both principal components.

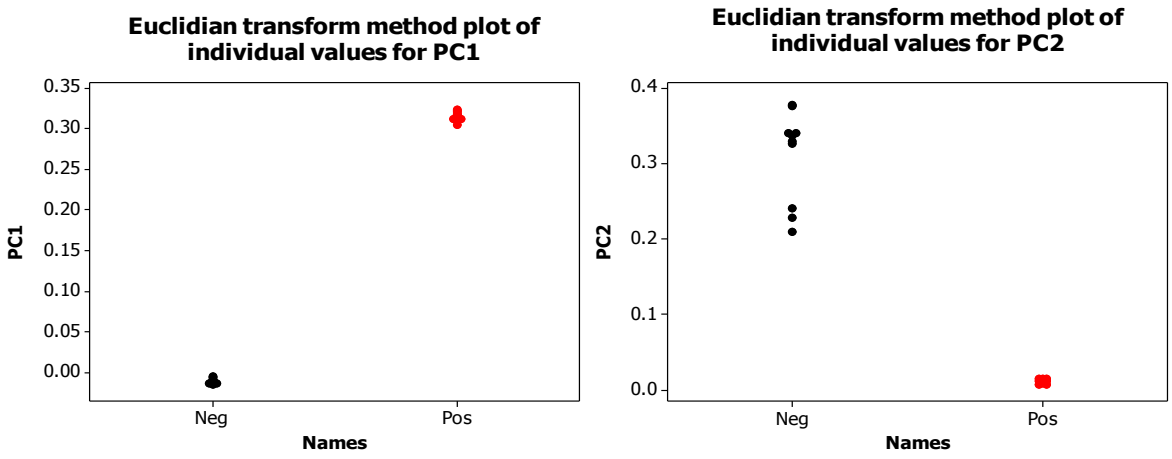


Figure 4-14 Plot for the individual values of the first and second principal components calculated from the data set after transformation using the Euclidian distance method. The names on the x axis are the channels from the differential mobility spectrometer.

Figure 4-15 shows the plots of the first two principal components against each other after transforming the data using both the summed method and the Euclidian distance method. Full separation can be seen in both data sets, but the data that has been transformed using the Euclidian distance method shows improved separation and grouping.

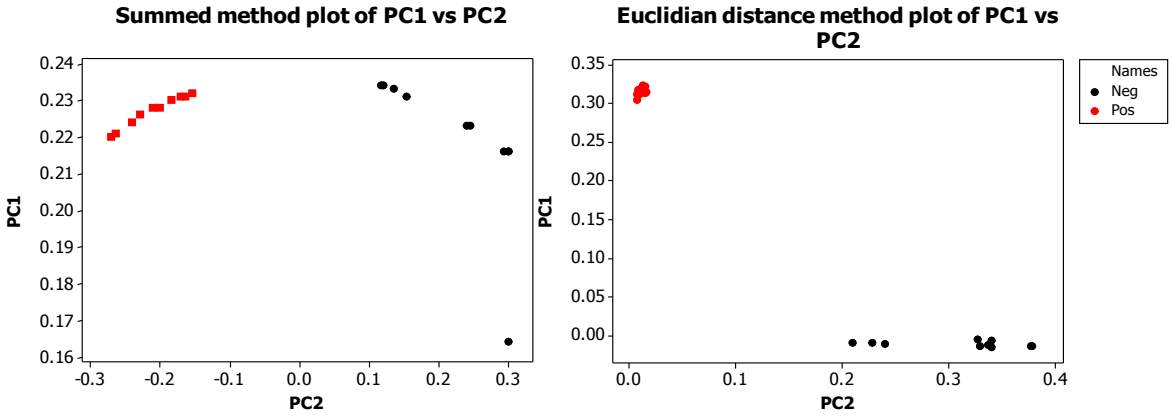


Figure 4-15 Plot of the first principal component values (on the y-axis) against the second principal component values (on the x-axis) calculated from both data sets after transformation using the summed method (left) and the Euclidian distance method (right).

4.3.5.3 Dispersion field sweep data sets

In day to day evaluation of data separating positive mode and negative mode data does not require PCA analysis for the position of the reactant ion peaks, the only signal in the data sets, is very reproducible.

The reproducibility of a plot is important in chemometric techniques for data analysis. In this instance these dispersion field sweeps are blanks, there is very little, if any, change in the position of the reactant ion peak. In contrast, features in non-blank data sets can drift, and consequently require complex pre-processing of the data sets before chemometric analysis. In this study such pre-processing was avoided to enhance the efficiency of the work and avoid any chance of favouring one or the other transform method.

Figure 4-13 shows the individual principal component plots for the first two principal components of the dispersion plot data set after transformation using the summed method. Although the separation of the positive and negative mode responses is a trivial exercise at one level it may be seen that separation is still not seen in the first principal component. In contrast, Figure 4-14 shows separation in both the first and second principal components if the Euclidian approach is adopted; see also Figure 4-15 that emphasises why this is the case. The Euclidean plots have retained their obvious differences and the principal component analysis has separated the groups, and generated precise clusters in the first principal component, with significantly improved discrimination when the first two components were plotted against each other.

4.4 Evaluation and testing the approach with modelled data.

Six tests were devised to test the Euclidian transform method. These six tests were based on the ways the peaks in a data set can change. The first three tests were designed to test that the method would work as the intensity of the peaks that differ between sets decreases. Two of the final three tests were designed to show that the method would work when only one peak differed between sets and when two peaks were unresolved with one set showing slightly more resolution. The final test was designed to show how the method performed when a peak within the data set drifted. To create the data files for these tests, a single data

file from the set described in 4.5.2 data set was selected and then modified to produce simulated responses in the following ways:

1) The average noise for the section -43 to -31.87 volts and 0 to 19.203 seconds was calculated for the selected file. Based on this average value the intensity at every point in the plot was changed in one of two ways:

2) A random number between zero and one and a half times the average value was added to every point to increase the amount of noise in the data set. Or,

2) The average noise signal was divided by four and subtracted from every point. A random number between zero and the divided average value was then added to every point. This method was used to adjust the noise in each sample while but increasing it. This method slightly reduces the average noise in the samples. This will be referred to as the adjusted noise method.

3) And 4) Gaussian peaks of differing values were created and inserted into the selected file to create new data sets with controlled differences. These peaks can be described in terms of compensation voltage location, retention time location, the spread of the peak in retention time and compensation voltage and the intensity of the peak. The compensation voltage location dictates where along the compensation voltage axis the peak maximum will be. The retention time location dictates where along the retention time axis the peak maximum will be. The spread of the peak in the compensation voltage and the retention time dictates the overall shape of the peak and the intensity dictates the maximum intensity which falls at the point where the compensation voltage location and the retention time location meet.

5) The file was cropped to exclude the RIP and areas where there were no peaks observed. The original file axes were -43 to 15 volts compensation voltage and 0 to 348.828 second and this was cropped to -9.61 to 7.97 volts and 79.875 to 280.063 seconds. This resulted in a surface of 31 points by 199 points.

These are the modifications that were made to every file created for the six tests. Figure 4-16 shows how the original file changed as these modifications were made. Details of the test data are included in Appendix I.

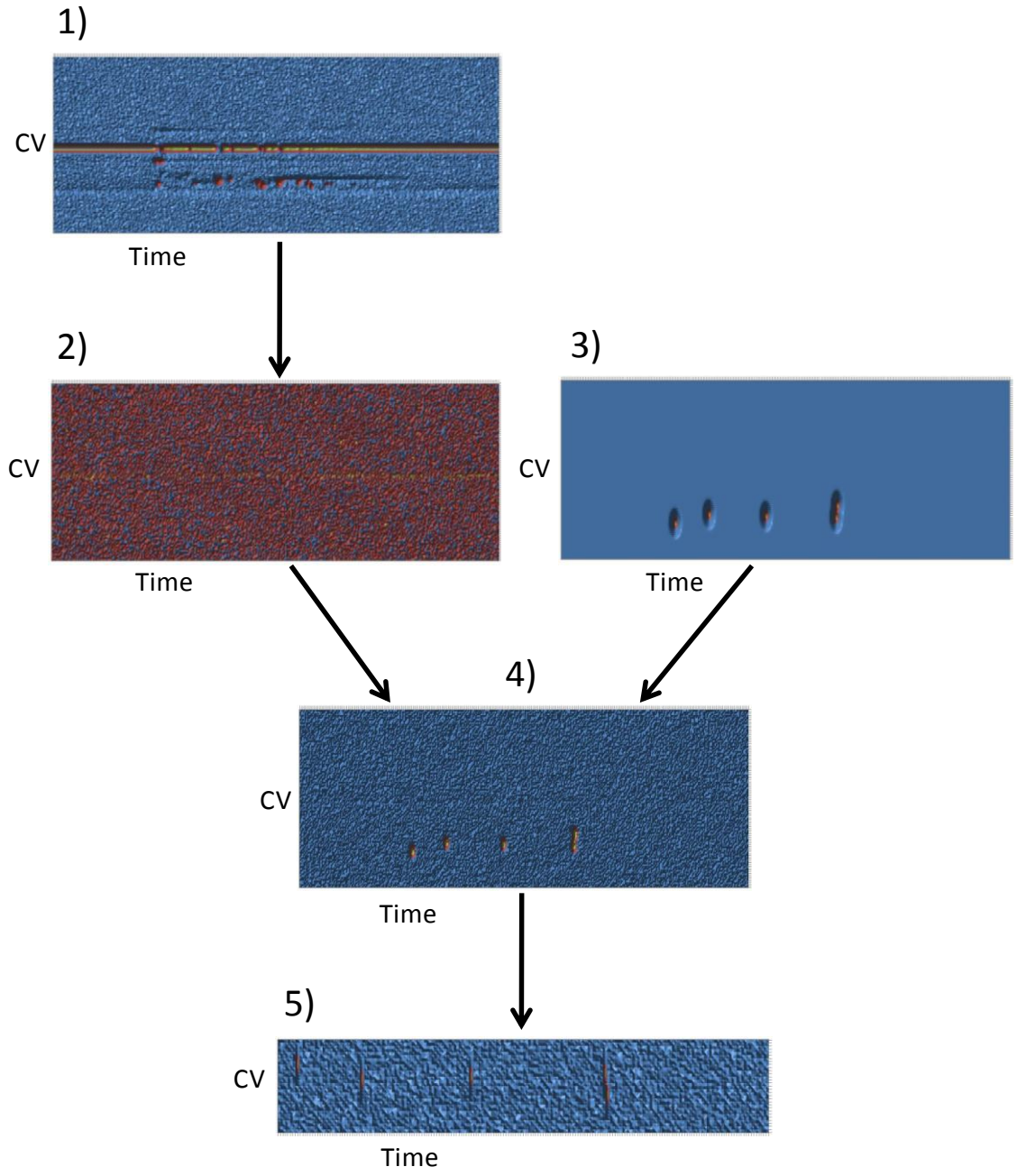


Figure 4-16 The DMS surface at different stages of modification. 1) Shows the original data file. 2) Shows the original file with the noise increased. 3) Shows the extra peaks (in this case five) that will be inserted into the file. 4) Shows the original data file with the noise increased and the new peaks inserted. 5) shows the original file with the noise increased and the new peaks inserted after it has been cropped to remove the RIP and the sections of the surface where no peaks appear.

The additional noise was recalculated every time a new file was created. This is because the same file is used as the base for each file created. If this was not done clustering of files in any analysis may have been artificially improved as the background noise pattern of every file would be the same.

For each of the six tests two sets of files were created, set A and set B, with the noise modified in one of the two ways described previously. Ten peaks, differing in retention time and compensation voltage, were inserted into every file.

The peaks inserted for each test were as follows:

4.4.1.1 Test 1 High intensity differences in half of the discriminant data.

Ten resolved peaks were generated into two sets of data (set A and set B). Each set of data contained noised replicates with five common peaks shared between the two sets and five peaks associated with the set These peaks had an approximate intensity of four times the maximum peak intensity of the original file. The noise was increased by 150 % of the average noise of the source file, see Figure 4-17. Where the intensity of the ten inserted peaks was approximately four times the intensity of the largest peak in the original data set. The noise was increased by one and a half time using the method described previously. The tables containing the details of the inserted peaks are Table 1 and Table 2 in Appendix I.

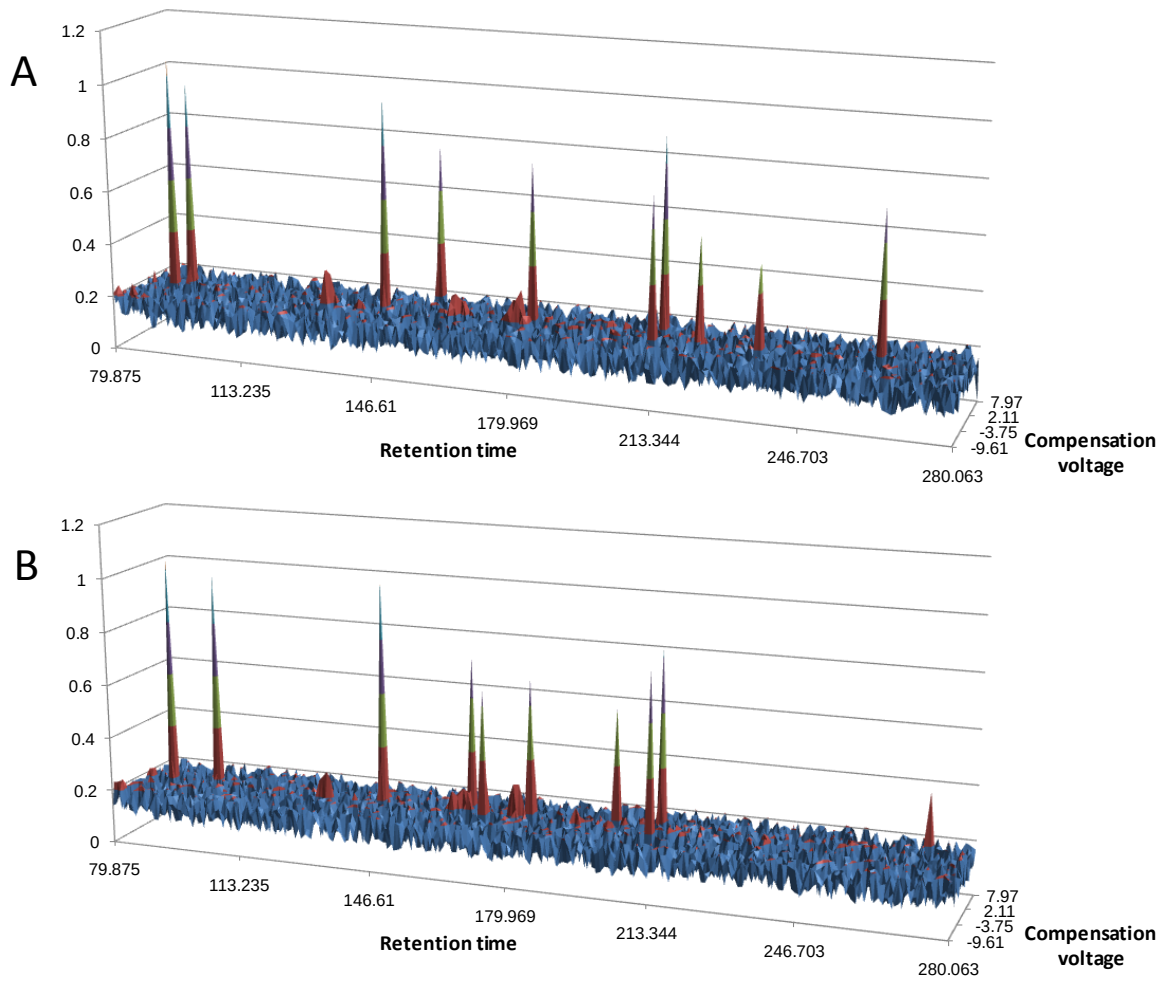


Figure 4-17 Plots of a single file from each set in Test 1. A is a plot of data file A1 and B is a plot of data file B7.

After the peaks had been inserted each data file was processed using the Euclidian transform method, and the two dimensional data transferred to Minitab 15 Statistical Software. A principal component analysis was performed extracting the first ten principal components. The first two principal components were then plotted against each other for all of the samples. This plot can be seen in Figure 4-18.

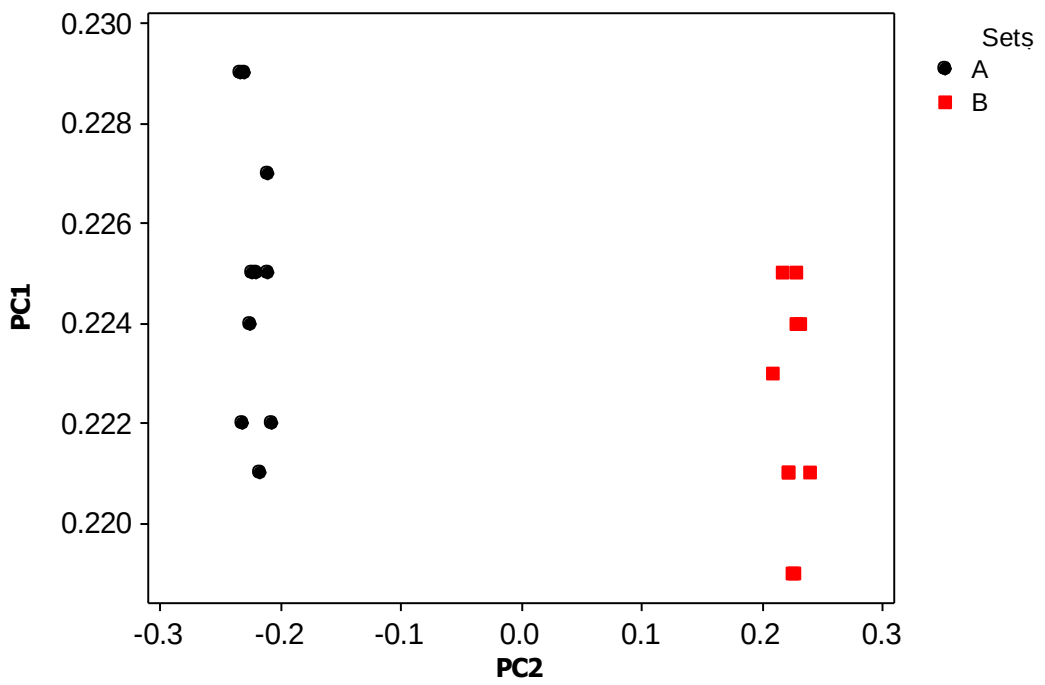


Figure 4-18 Plot of the first two principal components calculated for the transformed data files in Test 1. This shows a clear separation and grouping of the two sets, A and B.

A clear separation and grouping of the two sets, A and B, can be seen in Figure 4-18. This shows that the Euclidian transform method is able to differentiate between the two sets of data files described. This is not a rigorous test for the approach but provides a useful and encouraging starting point.

As an initial test five of the peaks in both sets were the same and the other five differed between sets. The results of the PCA show that it is possible to separate the two sets based on these differences. This is to be expected as the differences in the data sets can be visually distinguished. As highlighted by Figure 4-17. While this separation could have been done manually, for a larger number of data files per set, this could be a time consuming process, especially if the differences were not so obvious. This test has shown that the Euclidian transform method is a viable option for three dimensional data sets when a large number of differences are present.

4.4.1.2 Test 2 Medium intensity differences in a small proportion of the discriminant data.

The data sets were created in the same way as Test 1 but the intensity of the peaks was changed to between one and two times the maximum peak intensity of the original file and only two peaks differed in location, examples of the two sets are shown in Figure 4-19. The tables containing the details of the inserted peaks are Table 2 and Table 3 in Appendix I.

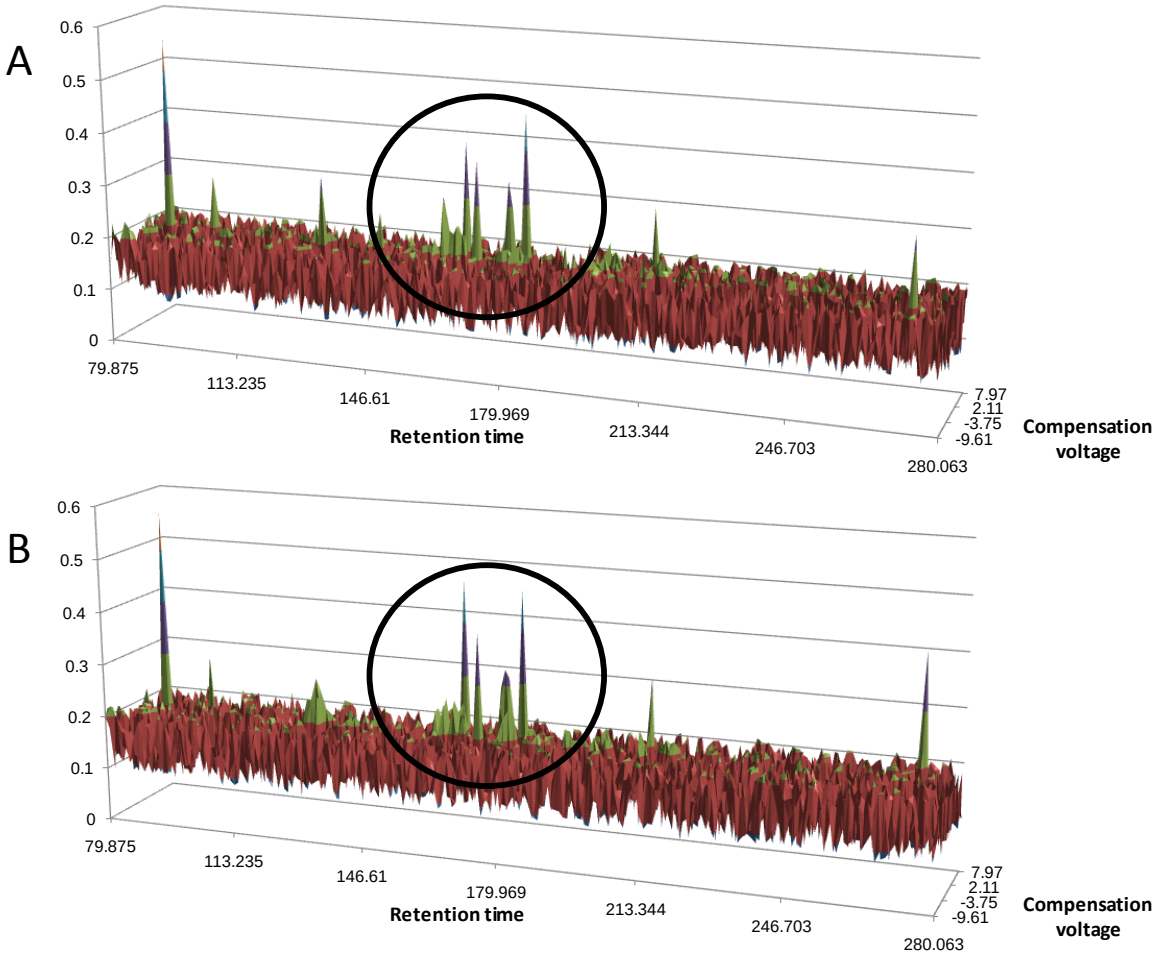


Figure 4-19 Plots of a single file from each set in Test 2. A is a plot of data file A8 and B is a plot of data file B5.

These data were processed and analysed as for Test 1 and the resultant plot of the first two principal components may be seen in Figure 4-18.

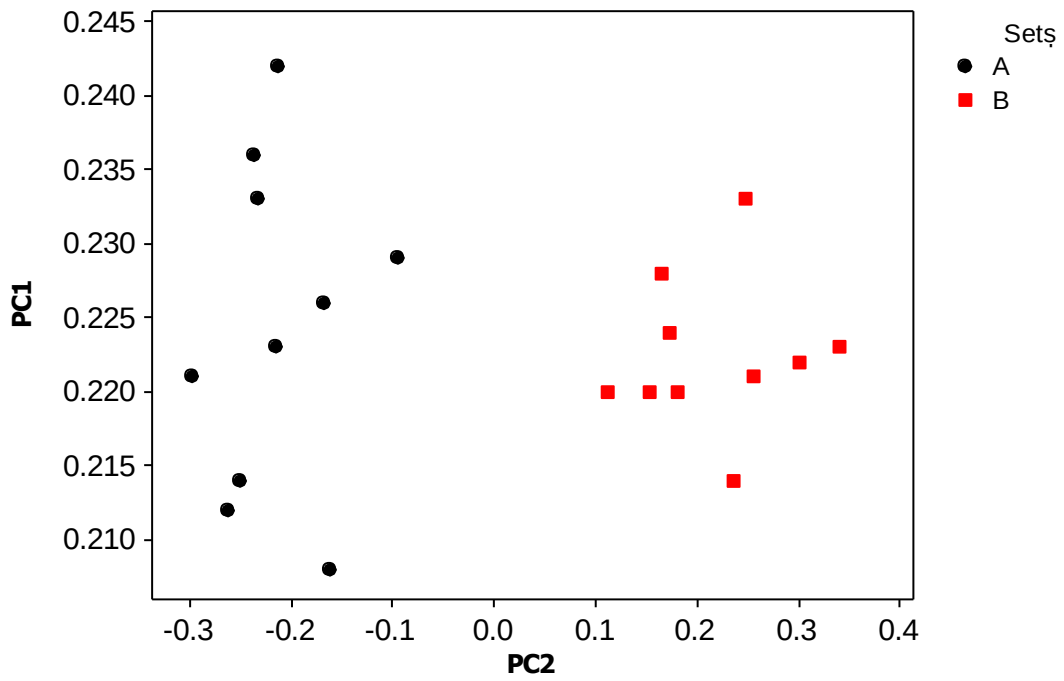


Figure 4-20 Plot of the first two principal components calculated for the transformed data files in Test 2. This shows a clear separation and grouping of the two sets, A and B.

The separation and grouping of the two sets, A and B, shows the effect of increasing the noise and decreasing the proportion of peaks that varied.

Test 2 was designed to test the Euclidian transform method when two sets of data that were similar to the sets in Test 1 but with the seeded peaks being a lower intensity were analysed. This represents sets that show large differences but not as great as Test 1 and therefore harder to distinguish. Again, the PCA was able to provide grouping and separation of the two sets based on the first two PCs, again showing that the Euclidian transform method is a viable option for data sets where there are large differences, but these differences are not as great as those tested in Test 1. A result of these differences being smaller is a wider spread within the groupings in the plot of the first two PCs, as show in Figure 4-20. As with Test 1, while the differences in Test 2 may be distinguishable by eye, for large, complex data sets an automated, multivariate analysis method is favourable and this shows that the Euclidian transform method is a viable option.

4.4.1.3 Test 3 Low intensity differences in a small proportion of the discriminant data.

The data sets were created in the same way as Test 2 but the intensity of the peaks was changed to less than one times the maximum peak intensity of the original file. In addition a third data set of ten surfaces was included in this test. An example of each set can be seen in Figure 4-21. Set A and set B contain 8 peaks that are the same and 2 peaks that are different. Set C contains ten peaks that are all different from set A and set B. The intensity of the inserted peaks was less than the intensity of the largest peak in the original data set. The noise was increased by one and a half time using the method described previously. The tables containing the details of the inserted peaks are Table 5, Table 6 and Table 7 in Appendix I.

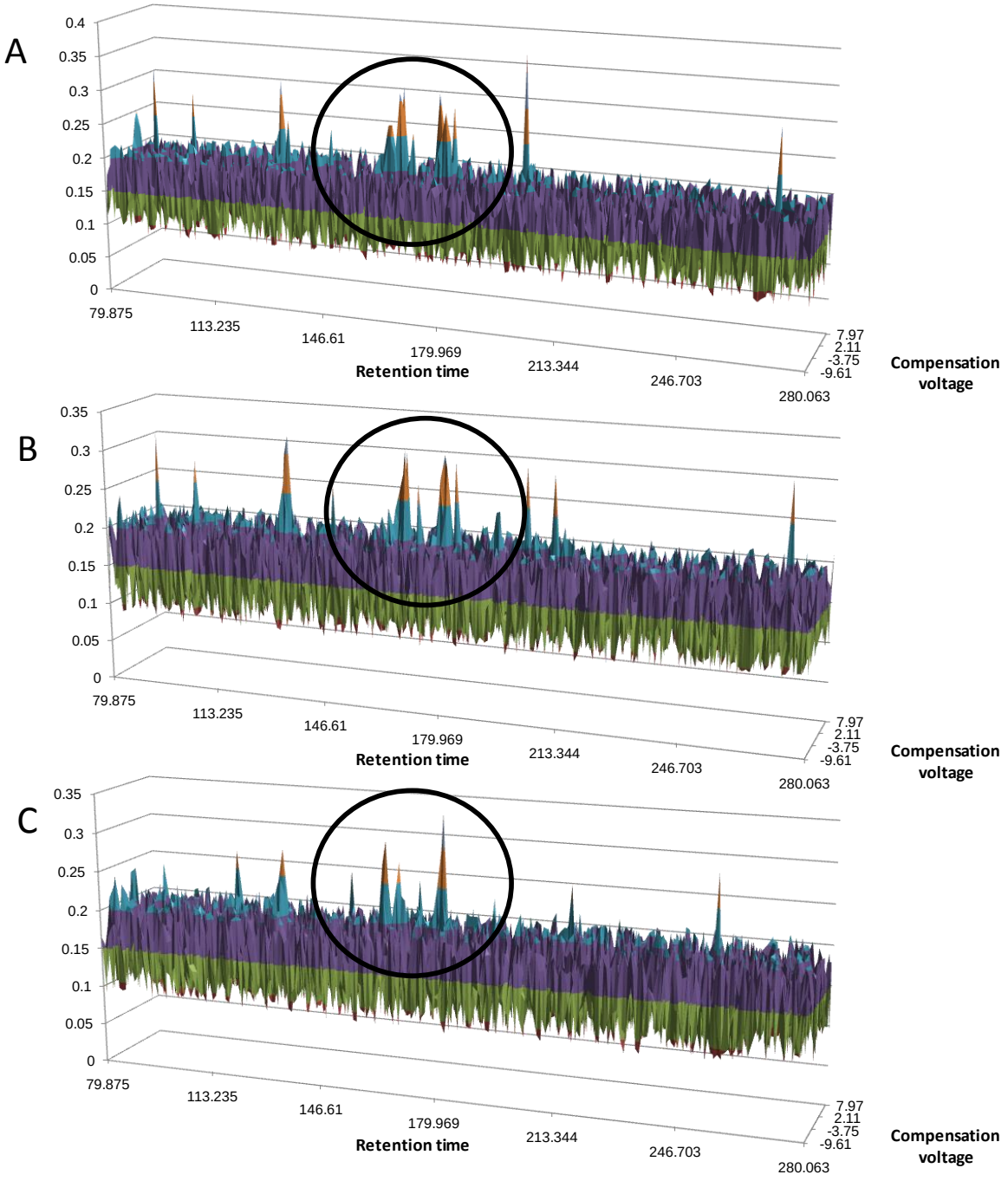


Figure 4-21 Plots of a single file from each set in Test 3. A is a plot of data file A4, B is a plot of data file B7 and C is a plot of data file C8.

Euclidian processing followed by PCA gave the result in Figure 4-22.

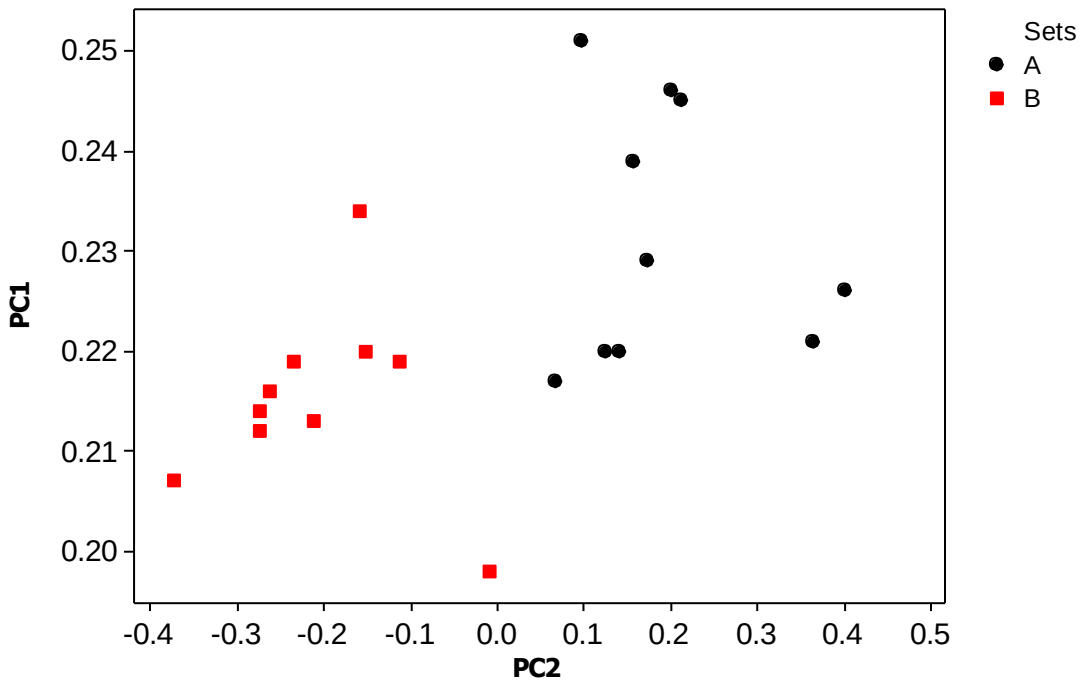


Figure 4-22 Plot of the first two principal components calculated for the transformed data files of set A and set B in Test 3. This shows separation of the two sets, A and B.

This was a more demanding test and the separation of sets A and B in the first two PCs was satisfactory albeit with greater scatter evident in the clusters. Comparison of these two sets with the third set required three PCs to be evaluated as can be seen in Figure 4-23.

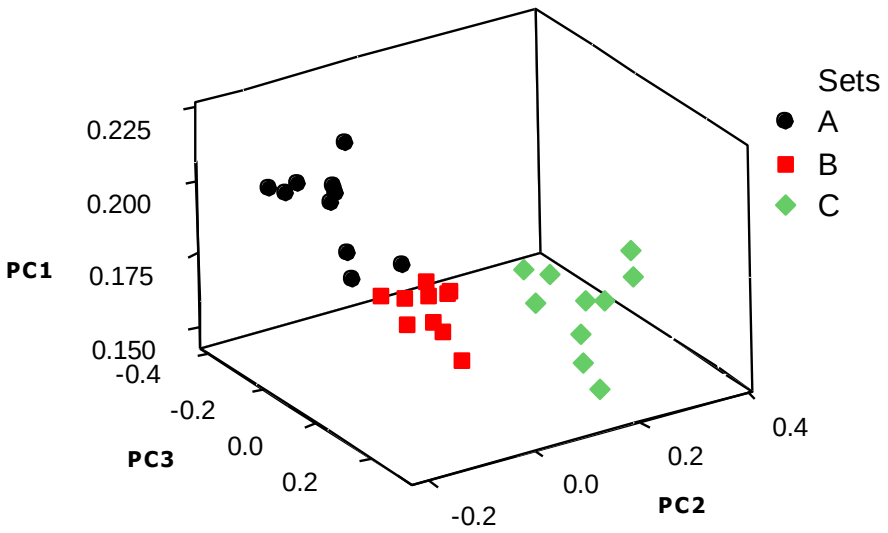


Figure 4-23 Plot of the first three principal components calculated for the transformed data files in Test 3. This shows separation of the three sets A, B and C.

Unsurprisingly Figure 4-23 shows the greatest separation between the third set and the first two sets; which had smaller differences between them. To show this more clearly a PCA was carried out on sets A and C. A plot of the first two PCs of this analysis can be seen in Figure 4-24.

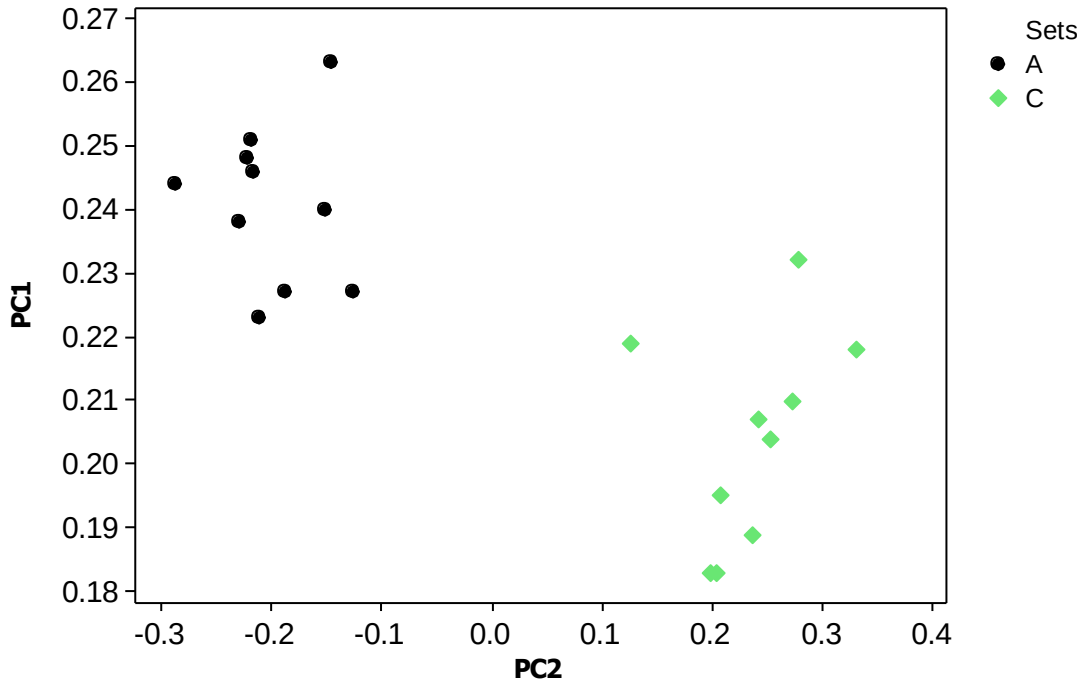


Figure 4-24 Plot of the first two principal components calculated for the transformed data files of set A and set C in Test 3. This shows separation of the two sets, A and C.

A much clearer separation between A and C can be seen when compared to Figure 4-22 which shows the separation for sets A and B. This is because the data files in set C are more clearly different to sets A and B than set A is to set B.

Test 3 was designed to test the Euclidian transform method when two sets of data files were given increased overall noise and seeded with peaks of a maximum intensity of less than that of the most intense peak in the original data file. This was to test whether the Euclidian transform method is a viable option for real data files as none of the seeded peaks would be more intense than those expected to be produced by an instrument. The PCA was able to give separate grouping of all three sets from each other as shown in Figure 4-23. As would be expected, set C is more separated from the other two sets than set A is from set B. As the differences between these sets are much less than the differences observed previously in

Test 1 and Test 2, the spread of the sets across the two PCs would be expected to be larger and the separation of the two groups less well defined and this is seen for sets A and B in Figure 4-22. With a greater number of differences, separation of sets A and C in a PC plot would be expected to be greater than the separation of sets A and B in Figure 4-22 and this can be seen in Figure 4-24. This shows that the separation seen between sets is reliant on the difference in the data files of the sets and is independent of the Euclidian transform method.

4.4.1.4 Test 4 Intensity difference in a single peak.

Ten resolved peaks were inserted into set A and set B. Nine of the peaks were identical in both sets. The tenth peak had the same retention time and compensation voltage location in both sets but a different intensity. The intensity of every peak inserted into the two sets was less than the maximum peak intensity of the original file. The noise was altered using the adjusted noise method as described previously and examples are shown in Figure 4-23. The tables containing the details of the inserted peaks are Table 8 and Table 9 in Appendix I. The results of the PCA after Euclidian processing are presented in Figure 4-26.

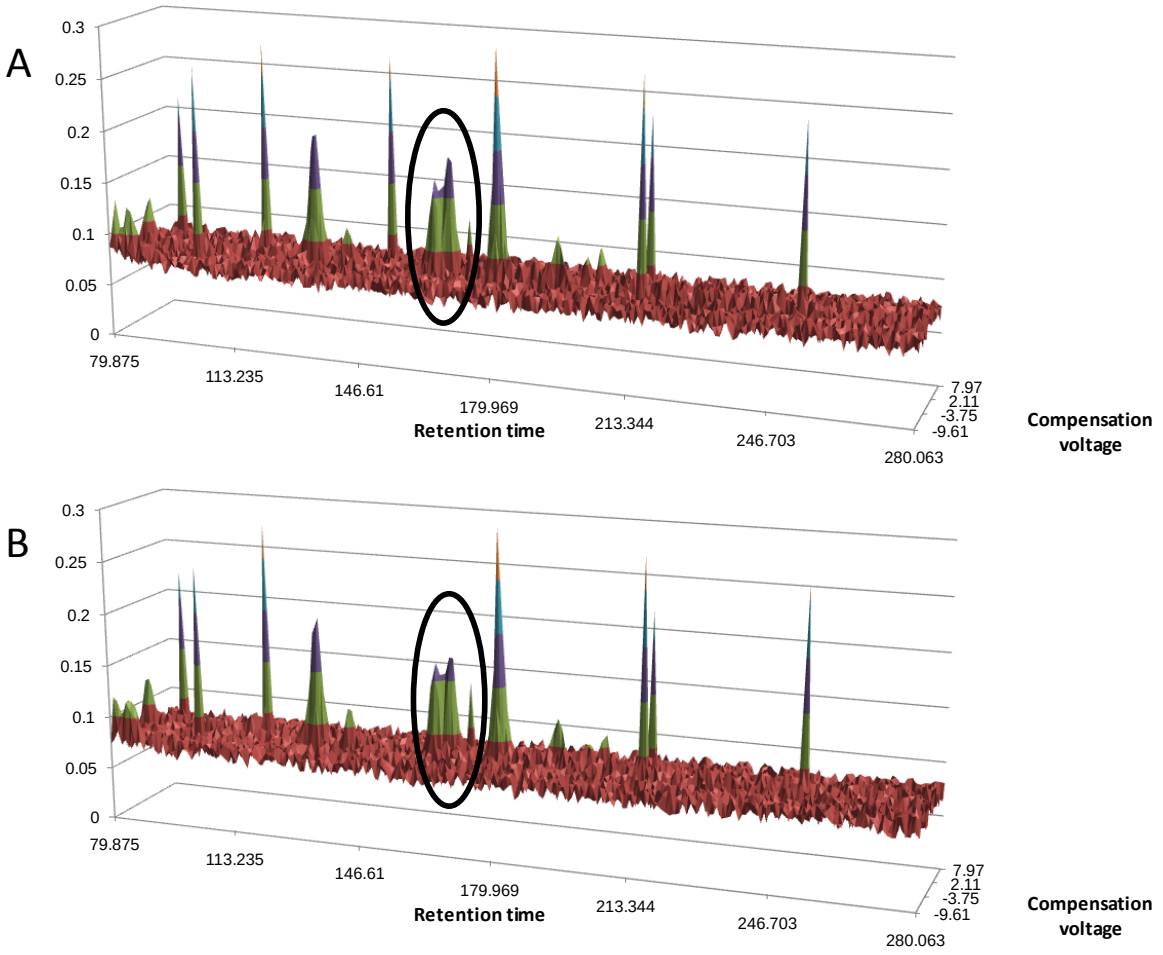


Figure 4-25 Plots of a single file from each set in Test 4. A is a plot of data file A5 and B is a plot of data file B4.

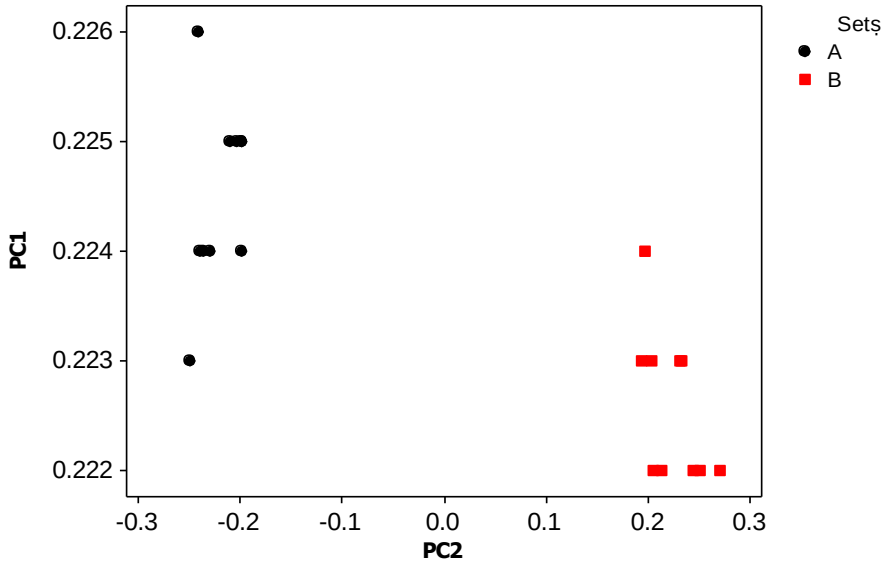


Figure 4-26 Plot of the first two principal components calculated for the transformed data files in Test 4. This shows a clear separation and grouping of the two sets, A and B.

A clear separation and grouping of the two sets, A and B, can be seen in Figure 4-26. Test 4 was designed to test the Euclidian transform method when a single peak changed in intensity in one of the two data sets. This test was designed to mimic the simplest of changes possible, such as a single biomarker in a diseased sample set that being slightly up-regulated compared to the controls. As Test 3 proved that the method was capable of distinguishing changes at an analytical level in data files with greatly increased noise, it was decided to use a different method to alter the noise, the adjusted noise method. Figure 4-26 shows the plot of the first two PCs from the PCA performed on the two sets created for Test 4. Separation can be seen between the two sets showing that the Euclidian transform method can be used to process data sets for analysis that have the smallest possible changes, a change in intensity of a single peak, and would be a viable method for research aimed at discovering biomarkers.

4.4.1.5 Test 5 Differences in resolution

Eight resolved peaks and two unresolved peaks were inserted into A and B. Eight of the peaks were identical across both sets, and two peaks were not fully resolved. Resolution is a measure of the separation of two overlapping peaks. The greater the separation between the two overlapping peaks the greater they are said to be resolved. The resolution of the two non-resolved peaks differed between the two sets with the peaks in set B being slightly more resolved than the peaks in set A. This was achieved by inserting the peaks in set B one unit further apart in the time axis. The noise was altered using the adjusted noise method as described previously. The tables containing the details of the inserted peaks are Table10 and Table 11 in Appendix I. Example data and the PCA analysis are presented in Figure 4-27 and Figure 4-28 respectively.

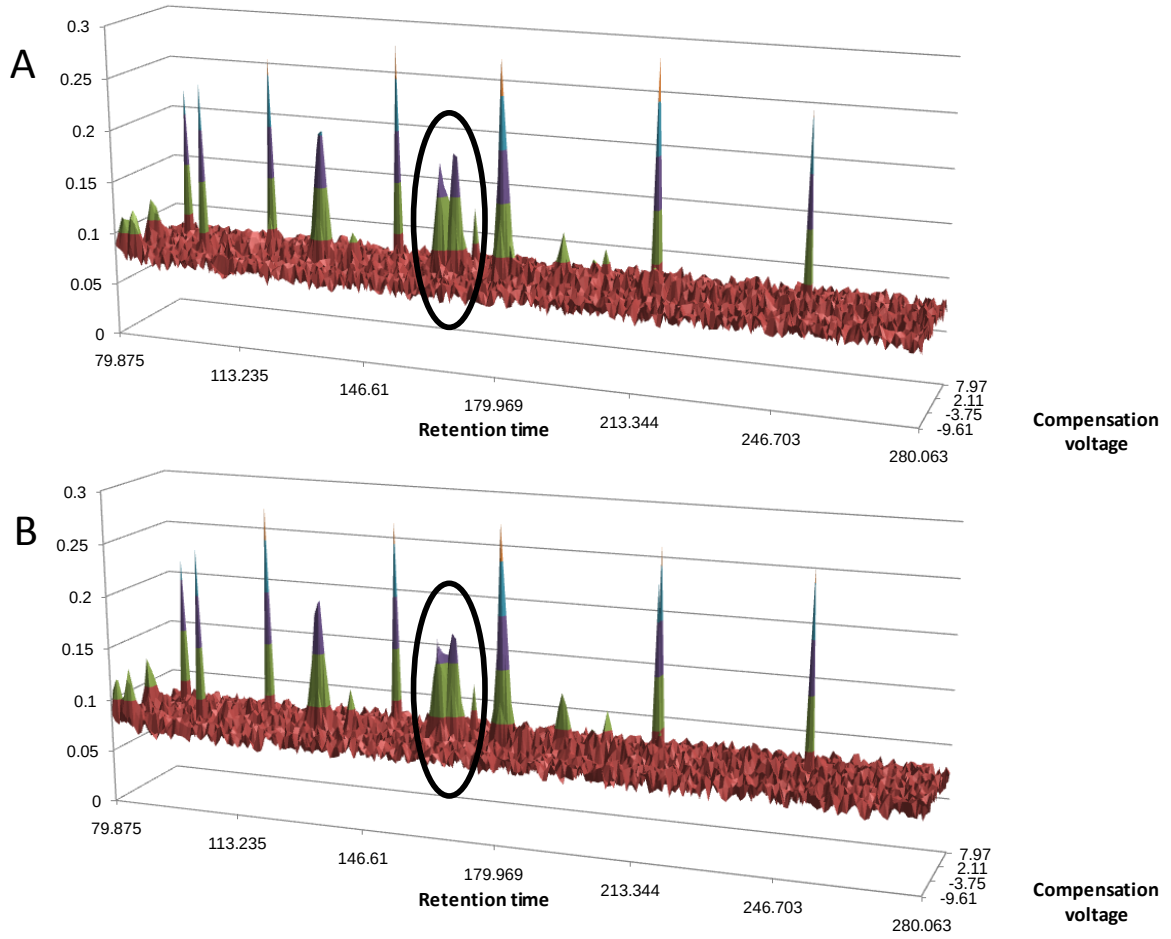


Figure 4-27 Plots of a single file from each set in Test 5. A is a plot of data file A3 and B is a plot of data file B6.

After the peaks had been inserted each data file was processed using the Euclidian transform method. After every file had been converted the resultant two dimensional data was transferred to Minitab 15 Statistical Software and a principal component analysis was performed extracting the first ten principal components. The first two principal components were then plotted against each other for all of the samples. This plot can be seem in Figure 4-28.

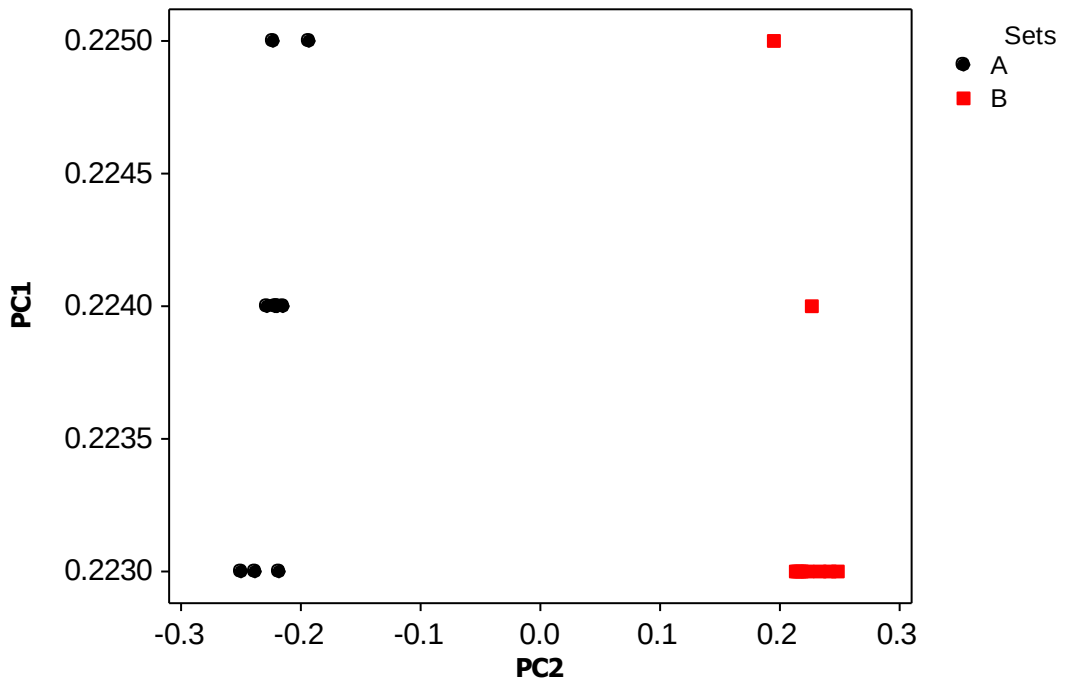


Figure 4-28 Plot of the first two principal components calculated for the transformed data files in Test 5. This shows a clear separation and grouping of the two sets, A and B.

Test 5 was designed to show that using the Euclidian distance method could be used to transform two data sets that contain two unresolved peaks at slightly different levels of resolution. This test is particularly important as it has been shown previously that current methods for transforming three dimensional data sets will not maintain the separation of peaks that fall at the same retention time but different compensation voltages. Figure 4-28 shows that when performing a PCA on two data sets where the only difference between the two is a different resolution of two unresolved peaks in the compensation voltage axis a clear separation between the two sets is seen in a plot of the first two PCs. This shows that, unlike the current methods, the Euclidian transform method can be used to convert data files from three dimensions to two dimensions and maintain the full information of the data file.

4.4.1.6 Test 6 Drifting phenomenon.

Ten resolved peaks were inserted into each file. Eight of the peaks were the same in both sets. In each set the compensation voltage and retention time of one peak, a different one for each set, was varied to mimic drifting. In the other set this peak did not drift. The tables

containing the details of the inserted peaks are Table 12, Table 13, Table 14 and Table 15 in Appendix I. The noise was altered using the adjusted noise method as described before. See Figure 4-29 and Figure 4-30 for examples of the data and the PCA results respectively.

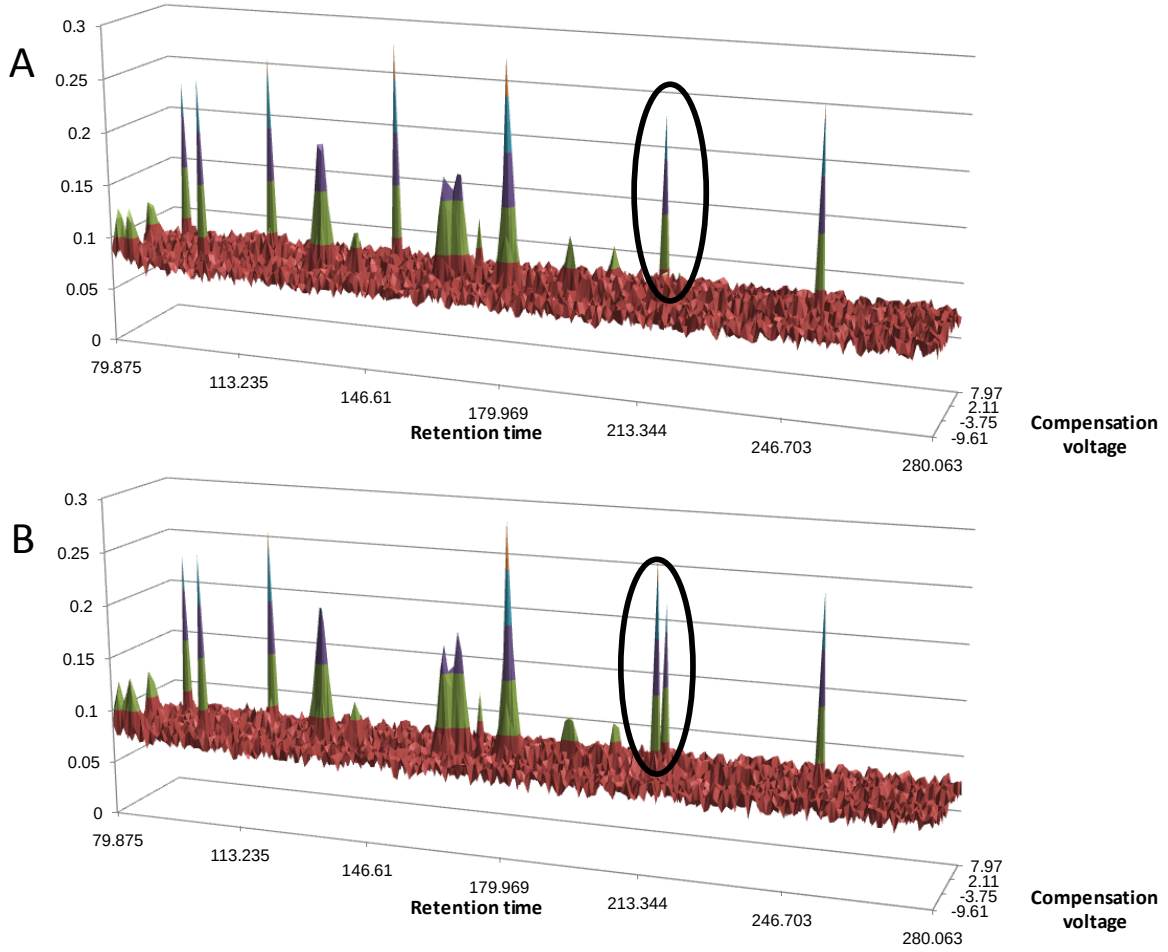


Figure 4-29 Plot of a single file from each set in Test 6. A is a plot of data file A3 and B is a plot of data file B10.

Figure 4-30 shows slight separation and grouping of the two sets. However, the spread of the data files in each set is considerably larger than before and some cross-over of the sets is seen.

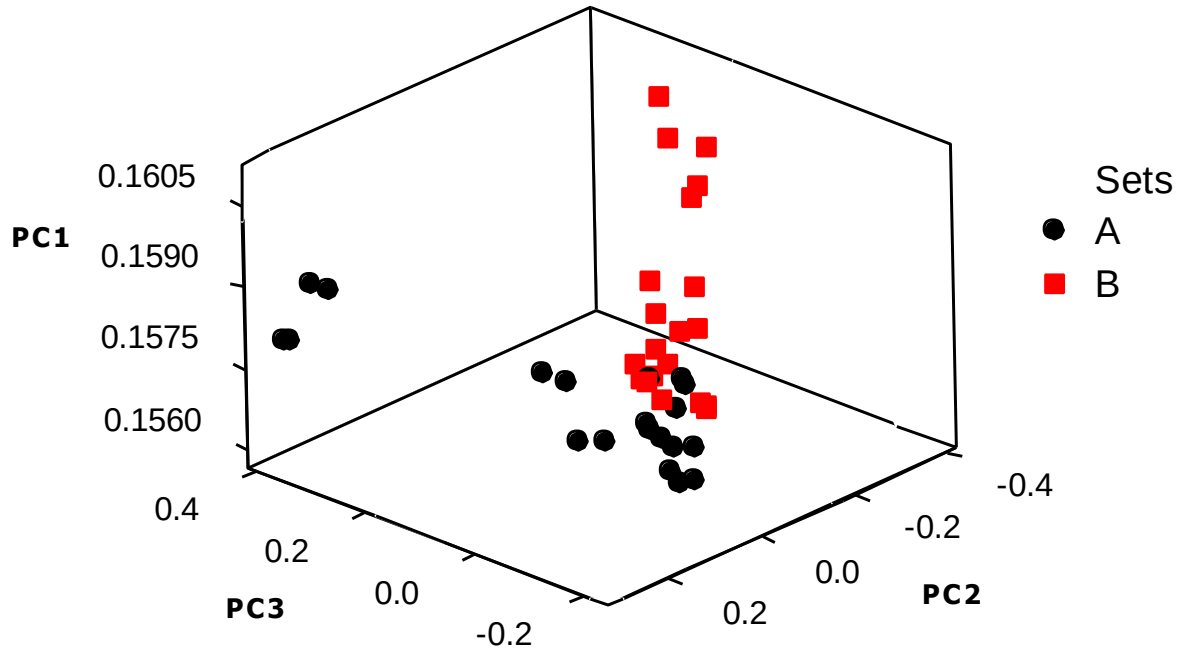


Figure 4-30 Plot of the first three principal components calculated for the transformed data files in Test 6. This shows slight separation and grouping of the two sets, A and B, but a wide spread in the set grouping with some cross-over.

Test 6 was designed to test how drifting peaks would affect the results of the PCA after the data files had been transformed using the Euclidian transform method. Drifting peaks is a problem in many areas of analysis, as can be exemplified by the large amount of research into and software available for aligning different types of analytical data. As the Euclidian transform method works on the basis of calculating a distance from the original of the data to every point on the DMS surface, slight changes in either compensation voltage or retention time could have an effect on the analysis. This was tested using two data sets that contained eight peaks with the same location and two peaks that were shifted to simulate drifting, with one peak drifting in each set. Figure 4-30 shows some grouping and separation of the two sets but there is cross-over of the two sets. This shows that alignment of the data is an important consideration and one that will still be needed with this approach.

4.4.2 Evaluation and testing with a spiked system.

A surrogate biogenic VOC profiles was generated using a complex mixture of VOCs, diesel headspace. The diesel standard was spiked with a single compound of known concentration at six different concentrations. These data files arose from another piece of research

involving the study of labels and impurities in fuel. The compound is an intelligent fuel tag. It was chosen to produce a known response, increase the sensitivity for specific analytes, in the samples. The exact nature and concentration of this compound was withheld (the data comes from another research project and so the nature of the compound is necessarily secret.) with them samples being split into 6 known groups. This provided a opportunity to test the new processing methods' ability, combined with the analytical systems used, to distinguish between the different sample sets. The compound was added at six different concentrations to produce six different data sets. These 6 concentrations provided 6 groups containing the following number of samples;

Table 4-3 Surrogate biogenic VOC profile groups along with sample numbers and known compound concentration

Data sets	Number of replicate analyses
A	5
B	4
C	5
D	4
E	5
F	7

All these samples were analysed randomly using dynamic headspace – gas chromatography – differential mobility spectrometry and Figure 4-31 is an example of data surfaces obtained. Three surfaces are shown, ranging, from left to right, from the lowest to the highest concentration of the added compound. The highlighted regions show some of the most visually obvious changes that occurred as the concentration of the added compound was increased. This provided a set of samples that mimicked the expected differences in breath samples from healthy and diseased participants and the changes investigated in the previous six tests described in 4.4.1.1 - 4.4.1.6, particularly the increasing concentration of particular peaks, which mimics the expected changes due to increased lipid peroxidation [49].

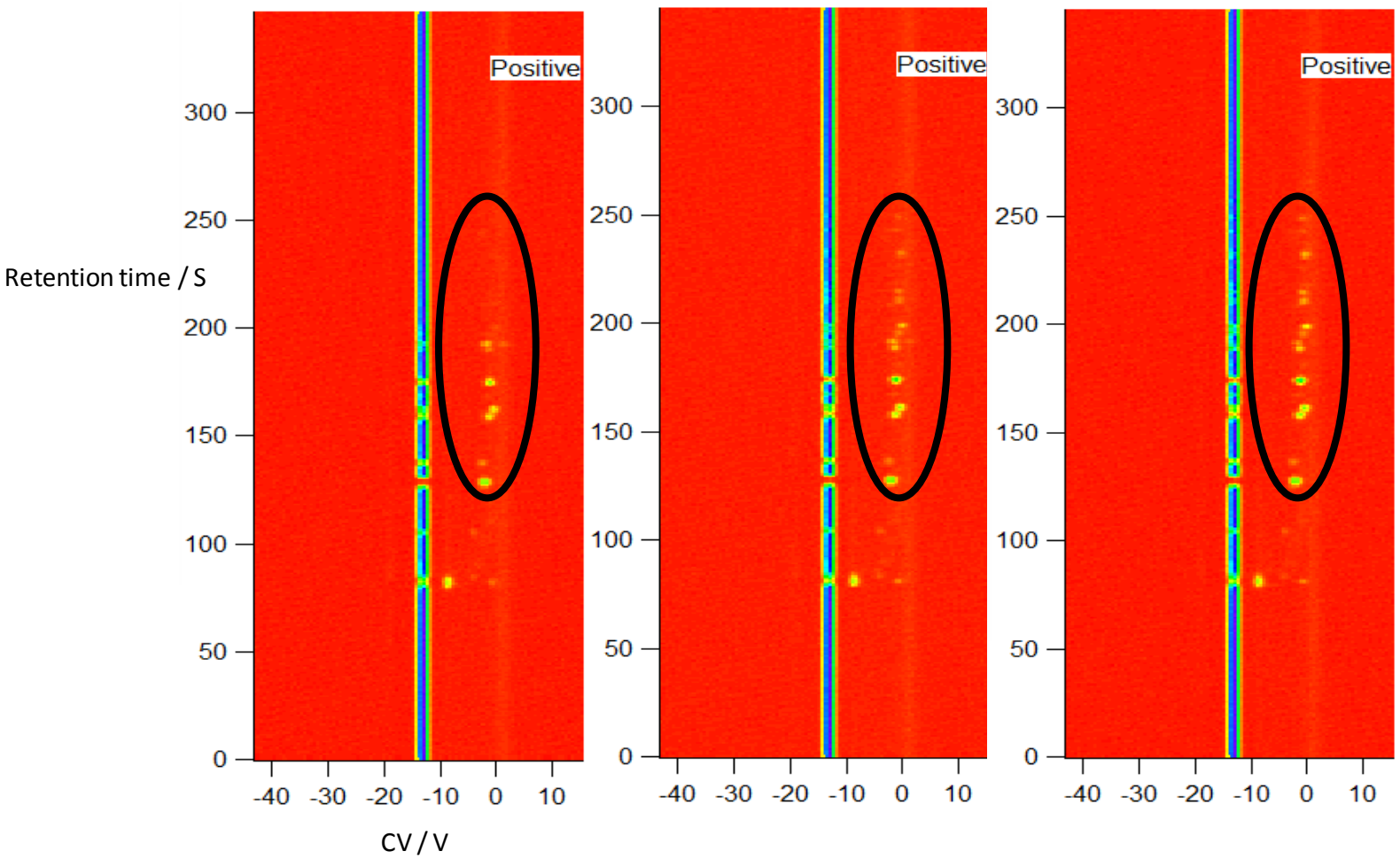


Figure 4-31 Three data surfaces generated from the positive DMS data. From left to right they show an increasing level of addition of the unknown compound. Highlighted is the region of the data surface that that shows the most change due to addition of the compound. While these are the most obvious visual changes others changes will be present which are not brought out by the colour shading.

This data set was baseline corrected and transformed from a three dimensional set to a two dimensional set using the summed method and the Euclidian distance method. 30 principal components were calculated for each data set. Initially plots of the principal components were examined to see if any separation of groups could be observed. As no separation was observed in the principal component plots, PC-DFA was performed on both data sets. The results of this analysis can be seen in Table 4-4.

The data processing was performed with Microsoft Excel 2007. The data surface was cropped and extraneous features such as labels and axes removed. The data were baseline shifted, important because the Euclidian transform programme determines the end of a row by looking for a zero in the data. Plots of the raw differential mobility data were produced in Igor Pro Version 5.0.1.0. A programme was written in Visual Basic 6.0 to perform the Euclidian distance transformation. The code for this programme is included in Appendix II. This programme was written to calculate the Euclidian distance for each point in the data series, order them, assign them a cell number and save the Excel file as a new, corrected file. Two dimensional plots of data after transform were plotted in Microsoft Excel 2007. Principle component analyses and plots of the principal components were carried out in Minitab ® 15.1.1.0. The principal component analysis (PCA) of each data set was carried out using a correlation matrix. In a correlation matrix each variable has a variance of one so that the total possible variance is equal to the number of variables. As the first components extracted contain the most variance, plotting these against each other can show grouping of similar data sets and so is a good method for analysing the data in this work. The maximum number of principal components that can be extracted is equal to the maximum number of variables. The analysis in this work was confined to the first few principal components as these are the ones most likely to contain the variance caused by real differences in the data sets.

Principal Component - Discriminant Function Analysis (PC-DFA) was used when no separation was seen in the principal component plots. PC-DFA is an *a priori* technique. This means that some knowledge of the samples is provided beforehand, in this case which set they belong to. With this knowledge the analysis seeks to minimise within set variance and maximise between set variance to assign the samples into their correct group. It uses the

PCs previously calculated to assign samples into sets. Increasing the number of PCs provided to the PC-DFA increases the amount of variance the PC-DFA has to differentiate between the sets. The more variance the PC-DFA has to use the more likely it is that it will be able to assign the data files to the correct group. Overtraining is a problem with PC-DFA and it is important to note the amount of variance being supplied to the PC-DFA as the more variance being used the more likely the data will be over-trained or assigned based on variation that does not represent real differences. The PC - DFA was carried out using a linear discriminant function on first seven principal components of the two dispersion field data sets analysed.

The PC - DFA was carried out using a linear discriminant function on the first seven principal components of the two dispersion field data sets analysed.

Table 4-4 The summary of the PC-DFA results from processing the data generated by applying the summed method of transforming.

Summary of PC-DFA classification using summed method data

	True Group					
Put into Group	A	B	C	D	E	F
A	5	0	0	0	0	0
B	0	4	0	0	0	0
C	0	0	5	0	0	0
D	0	0	0	4	1	0
E	0	0	0	0	4	1
F	0	0	0	0	0	6
Total N	5	4	5	4	5	7
N Correct	5	4	5	4	4	6
Proportion Correct	1	1	1	1	0.8	0.857
N = 30 N correct = 28			Proportion Correct = 0.933			

First 7 PC's used describing 97.3 per cent of the variance

Table 4-5 The summary of the PC-DFA results from processing the data generated by applying the Euclidian distance method of transforming.

Summary of PC-DFA classification using Euclidian distance method data

	True Group					
Put into Group	A	B	C	D	E	F
A	5	0	0	0	0	0
B	0	4	0	0	0	0
C	0	0	5	0	0	0
D	0	0	0	4	0	0
E	0	0	0	0	5	0
F	0	0	0	0	0	7
Total N	5	4	5	4	5	7
N Correct	5	4	5	4	5	7
Proportion correct	1	1	1	1	1	1
N = 30 N correct = 30	Proportion correct = 1					

First 7 PC's used describing 93.3 per cent of the variance

Table 4-4 shows the results of the principal component – discriminant function analysis results for the summed method transformed data set. Using the first seven principal components for the analysis, which describe 97.3 per cent of the variance, correctly assigns 90 per cent of the data sets to their correct group. Table 4-5 shows the principal component – discriminant function analysis results for the Euclidian distance transformed data set. Using seven principal components, the same number as used in the summed method analysis, which describe 93.3 per cent of the variance, correctly assigns 100 per cent of the data sets to their correct group.

4.5 Discussion

4.5.1 Comparison of the two methods when applied to a real data set

The first trial of a Euclidean approach was the analysis of dispersion data containing only the reactant ion peak described in 4.3.5. This simple preliminary test was chosen to test the data modelling approach on a trivial problem that nonetheless represented an increasing complexity over the five peak illustration, with the introduction of background noise into the system. Comparison of the outputs from the two transform techniques used was straightforward and interpretation of the responses would be.

Figure 4-9 and Figure 4-11 show the plots of the positive and negative channels after both transform methods have been applied to the original data set. The first thing to note is the size of the different outputs. The output from the summed method has one hundred data points, this is equal to the number of Rf steps there were in the original data set. Compared to the ten thousand data points seen when using the Euclidian distance method this is a very large loss of information. The size of the data set seen with the Euclidian distance method is dictated by the number of data points in the original set. That is to say, every data point in the original set is seen, nothing is combined. This means that while you see a single trend in the summed method data, an intensity decreasing as the intensity of the reactant ion peak decreases, the full data set is seen using the Euclidian distance method, with background noise being distinguished from signal.

In the plots for the Euclidian distance method the data corresponding to the reactant ion peak signal is contained mainly within the reassigned distance range four thousand to five thousand for the positive channel and two thousand to four thousand five hundred for the negative channel. This is due to the position of the reactant ion peak in the original data set and this peak is easily distinguished from the background noise signal. In the plots for the summed method the reactant ion peak cannot be distinguished from the background noise. In fact, without reference to the original data set it is very difficult to see where the reactant ion peak ceases to be present in the data. The only really discernable fact about the data from these plots is that the intensity of the signal due to the reactant ion peak decreases as

the Rf increases. The shift of the reactant ion peak, due to the increasing Rf, is not preserved using the summed method, whereas it is using the Euclidian distance method.

Preserving the intensity of the original signal is not possible using the summed method. This is shown in Figure 4-9 and Figure 4-11. The original data set has intensities ranging from approximately 0.1 to 0.35 v. using the summed method gives a summed intensity for each compensation field sweep which ranges between 12.35 and 11.7 V for the positive channel and 10.62 to 10.23 V for the negative channel in Figure 4-9 and Figure 4-11 respectively. This would make any quantitative analysis very difficult. In contrast, by using the Euclidian distance method quantitative analysis of the data is possible. Each distance can be traced back to the data point in the original data set. This means that it is possible to find every peak in the Euclidian distance data set that is due to a single component, and these intensities can be combined for quantitative analysis of the data.

The enlarged views in Figure 4-10 and Figure 4-12 show the full, rich data set that is available using the Euclidian distance method. As every point is plotted, the full profile of the reactant ion peak is observed. Due to the way the Euclidian distance method works, whole peaks are kept together. This is most easily seen in Figure 4-7, where the five peaks are fully coded in a small reassigned distance area, but is also seen in Figure 4-9 and Figure 4-11, where the reactant ion peak is kept within a finite region of the plot and maintains some of its' shape. This is the benefit of using a Euclidian distance method over purely unfolding the data. By maintaining the data structure in this way it is also easier to isolate and remove specific areas of a data set, such as the RIP, peaks or areas from of information. This would be much more complicated if the data was unfolded.

There are other benefits over just unfolding the data. This work covers combining retention time and compensation voltage but it is also possible to combine positive and negative data. This can be done by inverting one of the data sets and subtracting the intensity of each point to create a "difference surface" or by adding each matching point on the two surfaces to create an "additive surface". Another method to combine the two data sets would be to treat them as a continuous surface by combining them and then to read across both data sets for the transform in the same manner as for a single channel. This would maintain and

separate the structure of both data sets, although it would increase the file size and processing time.

As long as a common reference point can be found it should also be possible to include other information, such as mass to charge (m/z) ratio of the peak analysed in a mass spectrometer. This means peaks that occur at the same compensation voltage but have a different mass could be differentiated. The addition of underlying MS data could add a new level of complexity to the data and improve any analysis. It also eliminates the need for separate analysis of the MS and DMS analysis, both increasing the analytical power of the data and decreasing the amount of time and effort required to process it.

The high complexity of the data shown for the Euclidian distance method is due to two factors;

That the whole data set is transformed, including the large areas where nothing appears except background noise and,

That there has been no pre or post-processing of the data set.

The reason that no pre or post-processing was carried was so that any comparisons between the two techniques could be made directly, without fear that any pre or post-processing would bias the data towards either technique. A noise filter can be applied to the data either pre or post-processing, and this will reduce the complexity of the data set.

4.5.2 Surrogate biogenic VOC profiles

The next level of testing the Euclidian method involved a more complicated set of data and again the results were compared against analysis by a summation method. A set of highly complex mixtures were prepared to act as surrogate biogenic VOC samples. Into these mixtures a single component was added at six different concentrations to provide six different groups for the analysis. All of the data sets were visually checked for consistency before any processing was carried out, and the only pre-processing performed was baseline shifting of the raw data sets. This was to ensure the possibility of unintentionally introducing

bias into the data and method towards either approach was minimised. The assumption was performing the two alternative transforms on the same baseline corrected data would ensure that any differences seen in the outcomes would be representative of data processing outcomes.

After both transforms had been applied the first ten principal components were calculated for each data set. This was carried out in Minitab using The NIPALS method. The first four components were plotted against each other and singly to see if any separation could be seen. As no separation was seen it was decided to proceed to a principal component – discriminant function analysis of the data sets.

The principal component – discriminant function analysis was used as a method to test the differences between the two data sets. It was initially carried out using only the first principal component and then successive components were included in the analysis until the data sets for one method were fully assigned. This point was reached for the Euclidian distance transformed data using seven principal components. At this point the samples from all six groups were correctly assigned for the Euclidian distance set but only ninety per cent of the samples were correctly assigned for the summed method.

The variance contained within the seven principal components is significant. The less variance used to assigned samples to their respective groups means the more likely it is that the groups are being separated on real differences. The first seven principal components produced by the summed method contain ninety seven point three per cent of the variance whereas the seven principal components produced by the Euclidian distance method contain ninety three point three per cent of the variance. This means the assignment observed in the Euclidian distance set is more likely to be due to a real difference in the samples.

4.6 Conclusions

Several conclusions can be drawn from this work:

Applying a Euclidian transform method is a viable option for converting a three dimensional data set to a two dimensional data set.

Using the Euclidian transform method over other reported methods maintains all the information contained within the data set.

The Euclidian transform method has been shown to allow for better separation and grouping of data sets in a PCA when compared to the current methods used.

Initially it was necessary to show that the Euclidian-based method would transform a data set in a viable way, both maintaining the full information set but also maintaining the peak structure, something that purely unfolding the data cannot achieve. Once it was established that this method was viable a comparison of this method to the current standard of summing the data across the compensation voltage axis needed to be made to show that the Euclidian-based method was an improvement. After this was established using the dispersion field sweep data set, 4.3.5.3, a rigorous test of the method using data sets exhibiting the smallest likely differences between the samples was used to show the method was capable of supplying a data-set which could differentiate based on these changes. The method was then compared to the current standard again when applied to a real data set to show it was an improvement.

The main benefit of this method is its ability to maintain all the information contained within a sample. For complex data sets containing a large number of peaks a loss of fidelity, such as that which occurs when summing across the compensation voltage axis, may mask peaks of interest. The current situation is that the added benefit of the compensation voltage separation is lost. The technology is ahead of the data processing and this method aims to bring the data processing in-line with the current detection technology allowing researchers to utilise the full information contained in the data sets for advanced data processing methods such as PCA and PC-DFA.

Drifting of peaks is the next issue that needs to be tackled. 4.4.1.6 shows that this could be a major issue, although if the drifting can be limited by careful experimental control of the analytical instrument it may not present a huge problem. Alignment software exists for protein gels [156, 157] and similar system for DMS data may drastically improve the data analysis possible and should be an aim of those working in improving these processing techniques.

5 Design and implementation of a collaborative clinical campaign within the National Health Service (NHS)

5.1 Introduction

The development of a clinical breath sampling collaboration within the NHS, in this case Royal Preston Hospital, was key to this work as samples were to be collected within the NHS and analysed at Loughborough University. Developing this collaboration from the initial meeting with the clinical lead to the collection of the final samples and subsequent data processing, everything must be pre-planned and approved both by the clinicians whose knowledge of the workings of their NHS site are vital, but also by the appropriate ethics board.

To satisfy the ethics board several pieces of documentation are required and developing a process to ensure speedy delivery of these documents is vital to ensure that the research is not delayed. As well as satisfying all of the ethical requirements imposed upon any clinical research, there are several key parts of the research that must be discussed, such as breath sampling locations, the required services for sampling and the logistics of transferring the breath samples from the clinical site to the samples analysis site.

Breath sampling is a new field of research, and this means that it is unlikely that there will be a vast amount of experience both at the clinical site and at the samples analysis site. This means that education of both the clinical team as to what the requirements of the sampling are and conversely, education of the analytical team as to what can and cannot be expected of volunteers and clinicians within the NHS is important to allow for smooth completion of the research.

This chapter will describe the steps that were taken, the documents that were produced and the decisions that were made to ensure this research progressed in a timely fashion and any delays could be promptly dealt with.

5.2 Initial discussion with the Clinical lead

An initial discussion was held with the clinical lead at Royal Preston Hospital, Dr Stephen Fowler to discuss the research idea of taking breath samples from volunteers attending the two week wait clinic. This clinic is where NHS patients are referred by their general practitioner (GP) when a diagnosis of lung cancer is suspected. At this meeting the decision was made to collect samples from 40 volunteers, 20 diseased and 20 control. This was felt to be a sufficiently large cohort for this initial study. It was decided that patients would be recruited at their first outpatient visit. The initial inclusion criteria were decided upon and these were:

- Volunteer referred to the chest clinic with suspected lung cancer.
- Diseased – lung cancer confirmed
- Control – lung cancer excluded
- Age > 18 years

These were chosen to be as wide ranging as possible as any diagnostic test must be applicable to a wide range of subjects.

The exclusion criteria were also decided upon and these were:

- Co-morbidities that are likely to significantly affect metabolic status, i.e. diabetes mellitus, chronic liver failure, chronic renal failure.
- Other active malignancy

The co-morbidities were included as each produces a distinctive smell on the sufferers breath and as such it was felt this could significantly influence any breath analysis test.

5.3 Required documentation

There are several pieces of documentation required for a clinical study within the NHS. All of these were completed and submitted to the appropriate bodies.

5.3.1 Ethics documentation

Any study that involves interaction with NHS patients or their medical records requires ethical approval. For this work a research protocol was produced (Appendix VI) and this was submitted along with the online forms for ethical approval. The research protocol details the planned study and the justification for the research. This document outlined who was involved, the aims of the research, the methods to be used and the expected results. Patient information, consent and questionnaire forms were also submitted for approval. This was reviewed by the ethics board for the Lancashire Hospitals Trust at a review meeting that the researchers attended to answer any questions. On the basis of this documentation ethical approval was given for the study.

5.3.2 Researcher documentation

There are several documents required for researchers, both within and out with the NHS, based on the part of the work they will be carrying out. All researchers submitted a CV along with the ethics documents. For researchers external to the NHS interacting with participants a research passport was completed. This included a Criminal Records Bureau (CRB) check and an occupational health assessment. Once this research passport was completed an honorary contract was issued by Royal Preston Hospital allowing the external researcher to work unsupervised within the NHS.

5.4 Identification of sampling site

The requirements for the sampling site are detailed in Appendix VI. Two different sampling sites were identified based on these requirements with a third possible site also identified. The primary sampling site, an unused surgical suite in the Accident and Emergency department was used for the duration of the study, with the other two identified sites kept in case sampling had to be moved.

5.5 Transport logistics

The transportation of the adsorbent traps both to and from Royal Preston Hospital was carried out in two different ways. When a researcher was visiting from Loughborough University blanked tubes were carried with them at less than $0^{o}C$ in a polystyrene box on ice-packs. These were then transferred to a fridge until needed. Collected samples were

then returned to Loughborough University in the polystyrene box on ice-packs at less than 0^{o}C. When a researcher was not visiting all samples were sent via next day delivery on ice-packs at less than 0^{o}C in a polystyrene box to Loughborough University and blanked tubes were returned in the same fashion.

5.6 Installation and testing of breath sampling equipment

To make the breath sampler portable all of the components were stored in a wheeled suitcase and stored in a locked office. Once a sampling site had been identified based on the criteria set out in the research protocol Appendix VI, the sampling equipment was tested both mechanically and electronically to ensure it functioned correctly. Test samples were collected and analysed at Loughborough University to ensure the sampling equipment was functioning properly.

5.7 Training of hospital staff

To ensure that samples could be collected as and when participants were available, two members of hospital staff were trained to set up and operate the sampling equipment as well as train others. This meant that researchers from Loughborough University would not need to be present on every sampling occasion and that site at Royal Preston Hospital could be self-sustaining as new researchers could be trained to operate the sampling equipment.

5.8 Changes to the initial ethically approved study plan

During the study areas were identified where the protocol was lacking. The aim to recruit participants on one visit to the hospital and ask them to return for a second visit to give samples was found to be unsuccessful. To rectify this the protocol was changed to allow the usual 24 hour waiting period to be waived if the participant agreed so that samples could be collected during their initial visit to the chest clinic. This change was submitted to the ethics board and approved.

It was also decided that samples would be collected before diagnosis and only mention of general lung disease would be made in and patient information sheets with all references to lung cancer removed. As not every patient attending the two week wait chest clinic is aware that they may be suffering from lung cancer it was felt that an mention of the specific

disease would cause undue stress to participants, particularly as their Consultant would, most likely, not have discussed this with them yet. This change was submitted to the ethics board and approved.

5.9 Flow diagram and Gantt chart

During the course of planning and implementing this study two documents were produced to guide this, a flow diagram Figure 5-1 and a Gantt chart Figure 5-2. The flow diagram details all the steps that must be carried out to ensure that the study can start and the Gantt chart outlines time the main steps should be expected to take.

5.10 A guide to setting up a collaborative clinical study within the NHS

A more complete guide is given in Appendix VII. This guide is aimed at researchers with no previous experience in collaborative clinical studies and will guide them through each step and all of the documentation that is required as well as highlighting where possible delays may occur.

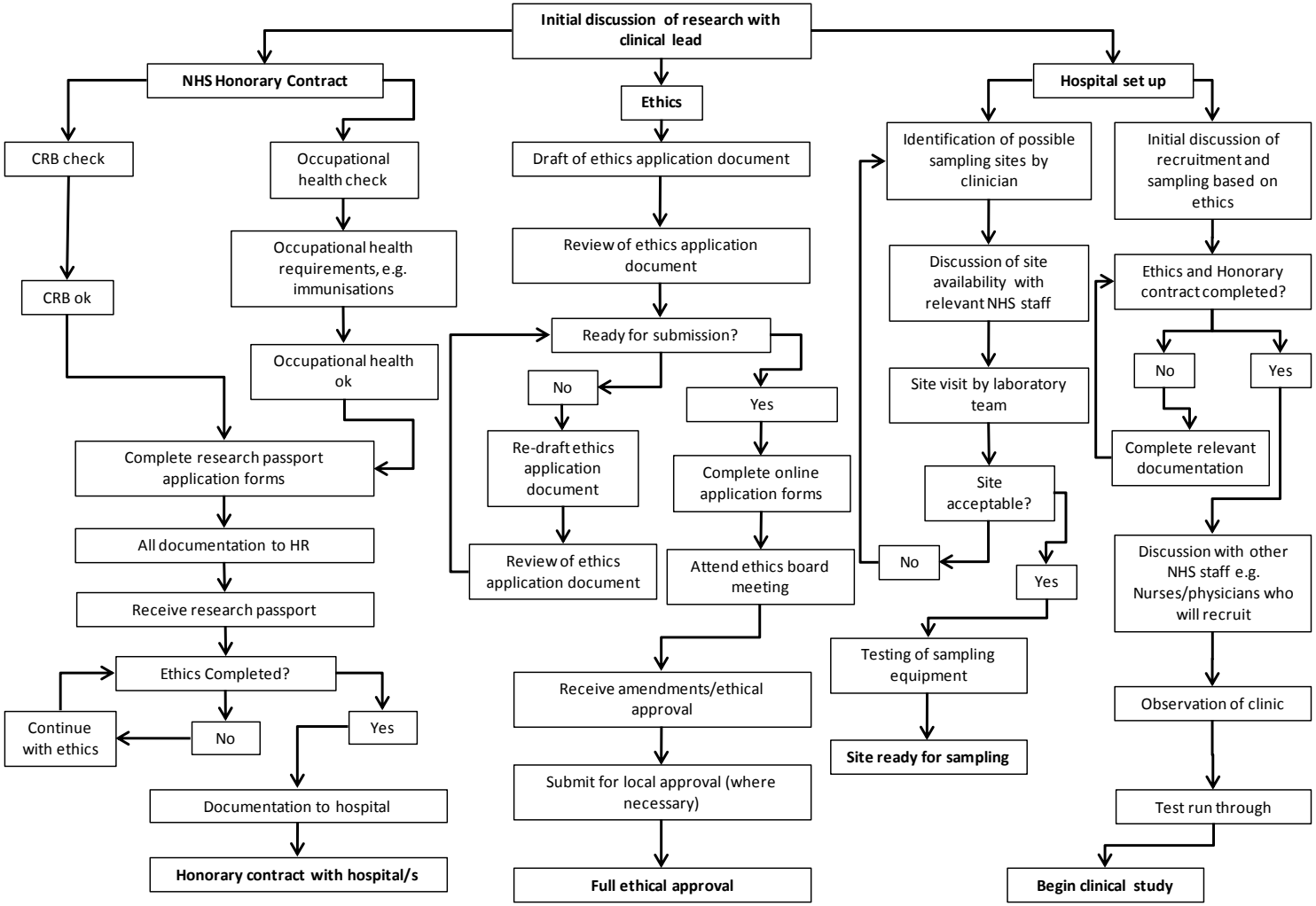


Figure 5-1 Flow diagram outlining the steps that need to be taken to implement a clinical campaign within the NHS as an external researcher.

ID	Task Name	Duration	Start	Finish
1	**Clinical campaign planning**	**107 days**	**Mon 15/02/10**	**Tue 13/07/10**
2	Intial discussion of reasearch	2 wks	Mon 15/02/10	Fri 26/02/10
3	**Ethics**	**90 days**	**Mon 01/03/10**	**Fri 02/07/10**
4	Protocol first draft	2 wks	Mon 01/03/10	Fri 12/03/10
5	Protocol final draft	2 wks	Mon 15/03/10	Fri 26/03/10
6	Online ethics forms	2 wks	Mon 29/03/10	Fri 09/04/10
7	submission of ethics	0 days	Fri 09/04/10	Fri 09/04/10
8	Ethics approval meeting	2 mons	Mon 12/04/10	Fri 04/06/10
9	Local ethics approval meeting	1 mon	Mon 07/06/10	Fri 02/07/10
10	Full ethical approval granted	0 days	Fri 02/07/10	Fri 02/07/10
11	**NHS research Passport**	**40 days**	**Mon 15/02/10**	**Fri 09/04/10**
12	Research passport documentation	1 day	Mon 15/02/10	Mon 15/02/10
13	Criminal Records Bureau check	2 mons	Mon 15/02/10	Fri 09/04/10
14	Occupational health check	1 wk	Mon 15/02/10	Fri 19/02/10
15	Issue of NHS passport	0 days	Fri 09/04/10	Fri 09/04/10
16	**Honorary contract**	**10 days**	**Mon 12/04/10**	**Fri 23/04/10**
17	submission of Research passport to NHS site	2 wks	Mon 12/04/10	Fri 23/04/10
18	Honorary contract issued	0 days	Fri 23/04/10	Fri 23/04/10
19	**Clinical location set up**	**19 days**	**Mon 26/04/10**	**Thu 20/05/10**
20	identification of possible sites	1 wk	Mon 26/04/10	Fri 30/04/10
21	discusion with relevant staff about site usage	2 wks	Mon 03/05/10	Fri 14/05/10
22	selection of site	0 days	Fri 14/05/10	Fri 14/05/10
23	testing of sampling set up at site	4 days	Mon 17/05/10	Thu 20/05/10
24	sampling site set up completed	0 days	Thu 20/05/10	Thu 20/05/10
25	**Recruitment and sampling run through**	**7 days**	**Mon 05/07/10**	**Tue 13/07/10**
26	discussion with clinical staff	1 day	Mon 05/07/10	Mon 05/07/10
27	Clinic observation day	1 day	Tue 06/07/10	Tue 06/07/10
28	test run through	1 wk	Wed 07/07/10	Tue 13/07/10
29	Begin clinical study	0 days	Tue 13/07/10	Tue 13/07/10

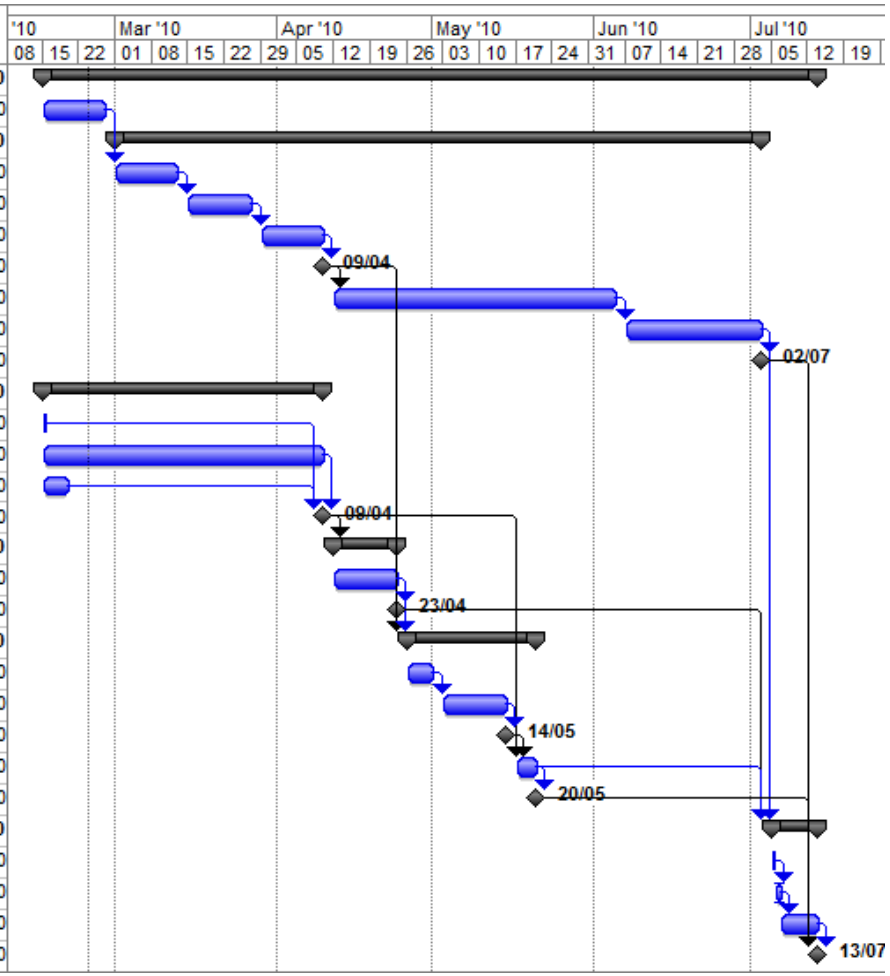


Figure 5-2 "Idealised" Gantt chart outlining the time each objective should take and when each objective should be started.

6 Exhaled Breath of Participants attending a chest clinic: a Pilot Study - An initial inspection and development of a workflow.

6.1 Introduction

VOCs in an individual's breath reflect the dynamics of their: environment, metabolism and, catabolism. Changes in exogenous factors, such as exposure to contaminants in inspired air [95, 161] may be controlled to some extent in a study. Dietary effects may be mitigated slightly by persuading participants not to eat or drink anything, other than plain water, for a period before sampling. Other criteria often require participants not to smoke. Keeping a record of diet in the week prior to sampling may be helpful in the subsequent analysis of data.

It is possible to identify non-biological exogenous components in breath profiles, based on complementary air sampling. Endogenous factors are more difficult to identify and control. For example, an individual's gut flora are difficult to regulate. Knowledge of the participants' recent bowel movements may help, but this assumes some homeostasis in human breath, not just intra-subject, but also inter-subject as it assumes gut flora will be the same across participants. At the moment it is sensible to record as much appropriate metadata as possible so any identified biomarkers can be evaluated against possible exogenous and endogenous contaminant sources. Such information will be invaluable in future comparative studies and one would hope meta-studies yet to be undertaken.

The variability of human breath makes searching for biomarkers of a particular disease problematic. It is not possible to purely group diseased and control subjects separately and compare them for differences, careful analysis of the meta-data is also required. One method of reducing the complexity of the task is to create a human breath database, containing all compounds that can be found in human breath. By carefully recording what compounds are found in both control and diseased subjects and the level they are found at, common compounds can be more easily discarded. While work exists that has detailed many compounds found in human breath there is no list that is continually updated and can be easily accessed by researchers from different groups. Similar lists in the fields of

metabolomics, such as METLIN [80] and The Human Metabolome Database (HMDB) [79], proteomics, such as Swiss-Prot [162], lipidomics, such as LIPID MAPS [163] and genomics, Such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [81], have shown to be an invaluable resource in these areas, and a database for compounds found in human breath would be very useful for breath analysis.

Certain information must be recorded with compounds, such as retention time, the diagnosis of the participant whose breath the compound was found in, the concentration the compound was found at and the m/z value for the most intense ions. This means that once a compound has been found, if it is again seen in another subject of a different diagnosis, this should be included in the database, meaning the knowledge base it provides is continually growing and being refined. As the database grew it would become easier for researchers to eliminate common compounds and focus on unique or uncommon markers that may be representative of a specific disease.

The first stage to establishing such a database is the collection, analysis, processing and logging of the compounds found in many breath samples from a variety of subjects. A complete clinical description of the participant is paramount, as the possibility of different diseases showing similar markers, particularly of oxidative stress, has already been shown by Phillips *et. al.* [36].

Obtaining the best possible chromatographic and mass spectrometric data from the instrumentation and careful data processing is required when assessing breath data. A single breath sample can contain many hundreds of compounds [40] many of which may not be resolved by the chromatography. The use of computer packages developed specifically to aid in the deconvolution of complex peaks such as AnalyzerPro from Spectralworks [164], Intellixtract from ACD labs [165] and the Automated Mass Spectral Deconvolution and Identification System (AMDIS) [166] from the National Institute of Standards and Technology (NIST) [167], makes this work possible. Where it is not possible to assign an exact compound, logging the most intense ions means that it can still be entered in the database with a tentative assignment and so can still be included in the knowledge base.

The aim of this study was to create a compound database by generating a list of every compound found in the breath samples collected from participants attending a chest clinic. This initial work outlines the difficulties associated with it and seeks to outline a workflow by which the data from this initial study can be processed.

The focus of this initial inspection is on the mass spectrometry data. This is because at the time of writing there were not enough samples available to do a preliminary analysis on the DMS data. As the mass spectrometry data can be processed as and when it is produced it provides an opportunity to assess the data during the study rather than waiting until the end.

6.2 Methods

6.2.1 Sample and information collection

Samples were collected on site at Royal Preston Hospital. The site was identified and set up in line with the process described and discussed in Chapter 5. Samples were collected from volunteer participants during their visit to a chest clinic run under the two week wait rule. To maintain age and sex matching only male participants over the age of sixty years old were used. Samples were categorised on the basis of the clinical description of the chest complaint the participants presented. Two broad categories were used; participants with a diagnosis of lung cancer were placed in a "Cancer" category and participants with a diagnosis that excluded lung cancer were placed in the "reference" group. The two week wait clinic is a fast-track clinical intervention that ensures that people with series signs and symptoms of respiratory disease are seen by a relevant specialist in the NHS within two weeks of their consultation with their general practitioner (GP)[a]; the full referral guidelines may be found on the National Institute for Clinical Excellence (NICE) website [168]. It is reasonable to assume that everyone attending such a clinic will have an underlying respiratory condition.

[a] While this is the current standard is should be noted that this may change depending on the political landscape

All samples were collected using the adaptive breath sampler described in chapter 2 according to the protocol which was granted ethical approval by the Lancashire Hospital Trust Regional Ethics Board and the Royal Preston Hospital local ethics board. Each participant provided 2 samples, and another sample was taken sample from the medical air line. Medical air was supplied by the hospital air system at four bar pressure. After sampling the volunteer was asked to fill in a questionnaire (Appendix V) aimed to highlight any factors, both exogenous and endogenous, that may influence the analysis.

6.2.2 Sample transport and storage

Samples were transported in two different ways. When a researcher based in the laboratory was present the samples were couriered back by the researcher in a sealed container maintained at less than four degrees centigrade by re-freezable ice packs. When a researcher from the laboratory was not present during sampling, the samples were sent through a courier service for next day delivery in a sealed container maintained at less than four degrees centigrade by re-freezable ice packs. Once samples arrived at the laboratory they were stored at four degrees centigrade in a dedicated LEC LR207 spark-free laboratory fridge.

6.2.3 Sample selection

Although two samples were collected from each volunteer instrument faults sometimes interfered with the subsequent analysis. The current sample list, along with the relevant clinical data are summarised in Table 6-1.

Table 6-1 Sample list detailing the volunteer code, sample I.D.'s, date collected and the volunteers' diagnosis.

Volunteer code	Sample I.D.		Date collected	Diagnosis
CV-05	T270OC1	T580OC2	15/10/2009	Background emphysema, no evidence of malignancy
CV-06	T129MF1	T128MF2	15/10/2009	No malignancy
CV-07	T182FS1	T190FS2	21/10/2009	Currently unknown
CV-08	T129BC1	T281BC2	04/11/2009	Appearances consistent with a primary pulmonary neoplasm with extensive nodal disease. StagingT4 N2/3 M1/0. Squamous cell carcinoma
CV-09	T288PS1	T287PS2	26/11/2009	Currently Unknown
CV-10	T182CH1	N/A	26/11/2009	No malignancy
CV-11	T572RW1	T141RW2	18/12/2009	Possible bronchiectasis, no malignancy
CV-12	T067FH1	T065FH2	03/02/2010	No malignancy
CV-13	T186MR1	N/A	3/02/2010	Provisional radiological diagnosis t4, n3, m1. Small cell lung cancer.
CV-14	T184JL1	T282JL2	18/02/2010	Appearances of bronchial carcinoma if histologically proven then staging would be T4, N2, M0
CV-15	T123PH1	N/A	18/02/2010	Benign TB a possibility, no malignancy
CV-16	T487RA1	T283RA2	11/03/2010	Appearances are those of a primary pulmonary neoplasm at the right pulmonary hilum. Radiological staging= T4, N2, M0
CV-17	T286KK1	T139KK2	11/03/2010	Appearances suggestive of a primary pulmonary neoplasm. Provisional stage= T4 N3 M0. Non-small cell carcinoma

At the time of writing a full set of samples had not been obtained. For this preliminary evaluation of the data and development of a work-flow only samples from participants who had been diagnosed with lung cancer were inspected.

6.2.4 Sample analysis

Sample analysis was randomised for samples from volunteers taken on the same day. Due to the restriction placed on analysis that all samples should be run within forty eight hours of sampling and the fact that sampling occurred weekly, if only one volunteer was sampled, samples could not be randomised.

The samples were analysed using the dual detector instrument described in Chapter 3. A system blank was run before and in between any sample analysis began to verify that the system was free from of contaminants and functioning. The instrument settings used were the same as previously described in chapter 3. The GC temperature program used is shown in Figure 6-1.

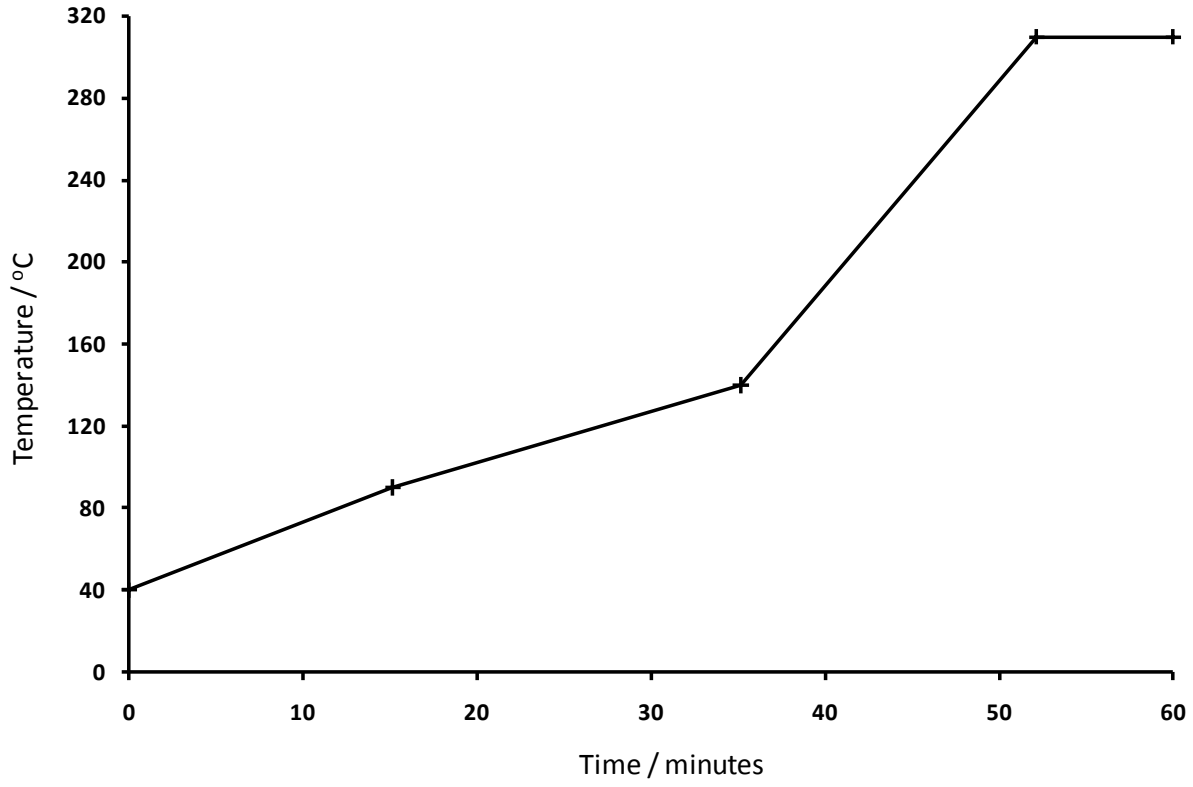


Figure 6-1 GC temperature program used for the analysis of breath samples.

This instrument produced two different sets of data, one from the mass spectrometer and one from the differential mobility spectrometer. Both sets of data were visually checked before any data processing was carried out.

6.2.5 Data processing

Once the samples had been analysed the mass spectrometry data was analysed in three different stages, these are:

1. Visualisation,

2. Peak examination and,

3. Identification and detected volatile table generation.

6.2.5.1 Visualisation

Each data set was selected in turn and visually inspected using Varian MS Workstation MS Data Review, Version 6.9 (Service Pack 1). Samples where the analysis had introduced artefacts or the instrument had not operated within specification were discarded. Samples that were contaminated with high levels of background volatiles were discarded. The sample was then inspected for resolved chromatographic peaks. Fully resolved peaks were identified visually and by examination of the underlying mass spectrum. Figure 6-2 shows an example of how peaks were visually distinguished as either resolved or unresolved.

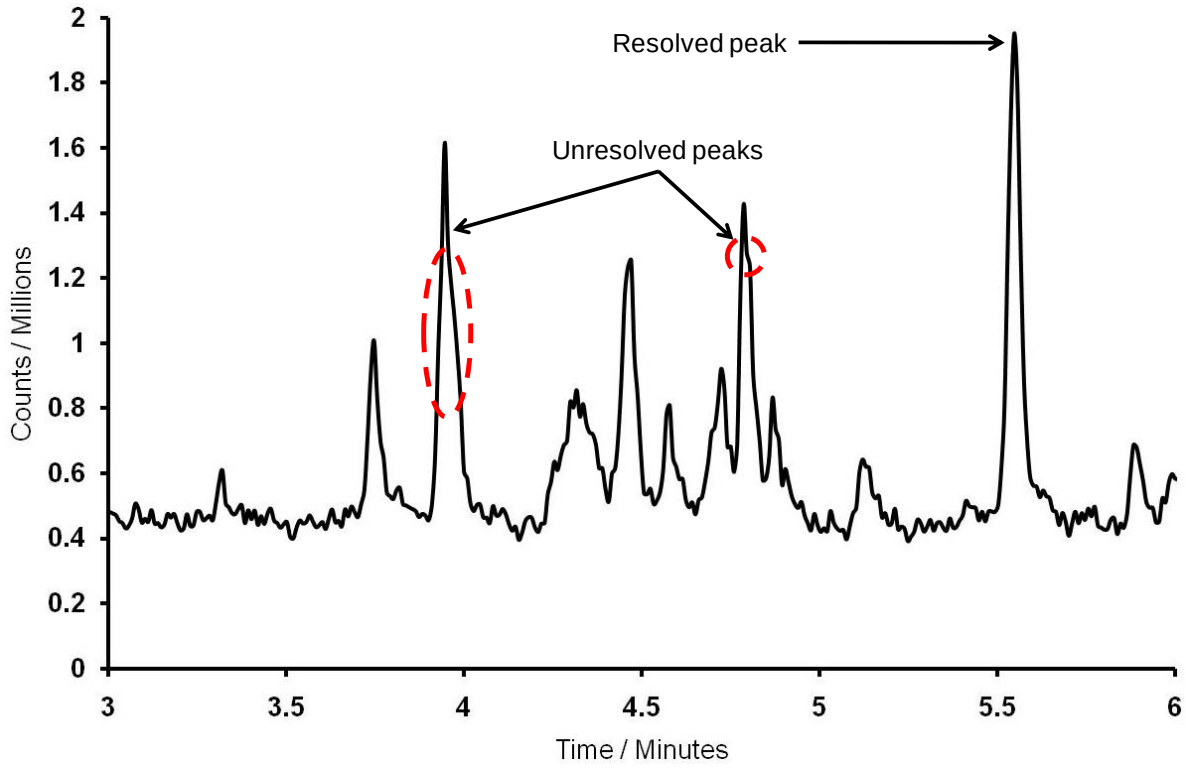


Figure 6-2 Example chromatogram showing both resolved and unresolved peaks. The 'shoulder' that appears on the peaks, highlighted by the red dashed circles, are due to ions from different components that have not been fully separated while passing through the chromatography column.

If a peak did not obviously fall into one or other groups, the underlying mass spectra were assessed to evaluate the peak homogeneity. This was done by examining the mass spectra at different points in the peak and assessing if the ions present changed, see Figure 6-3.

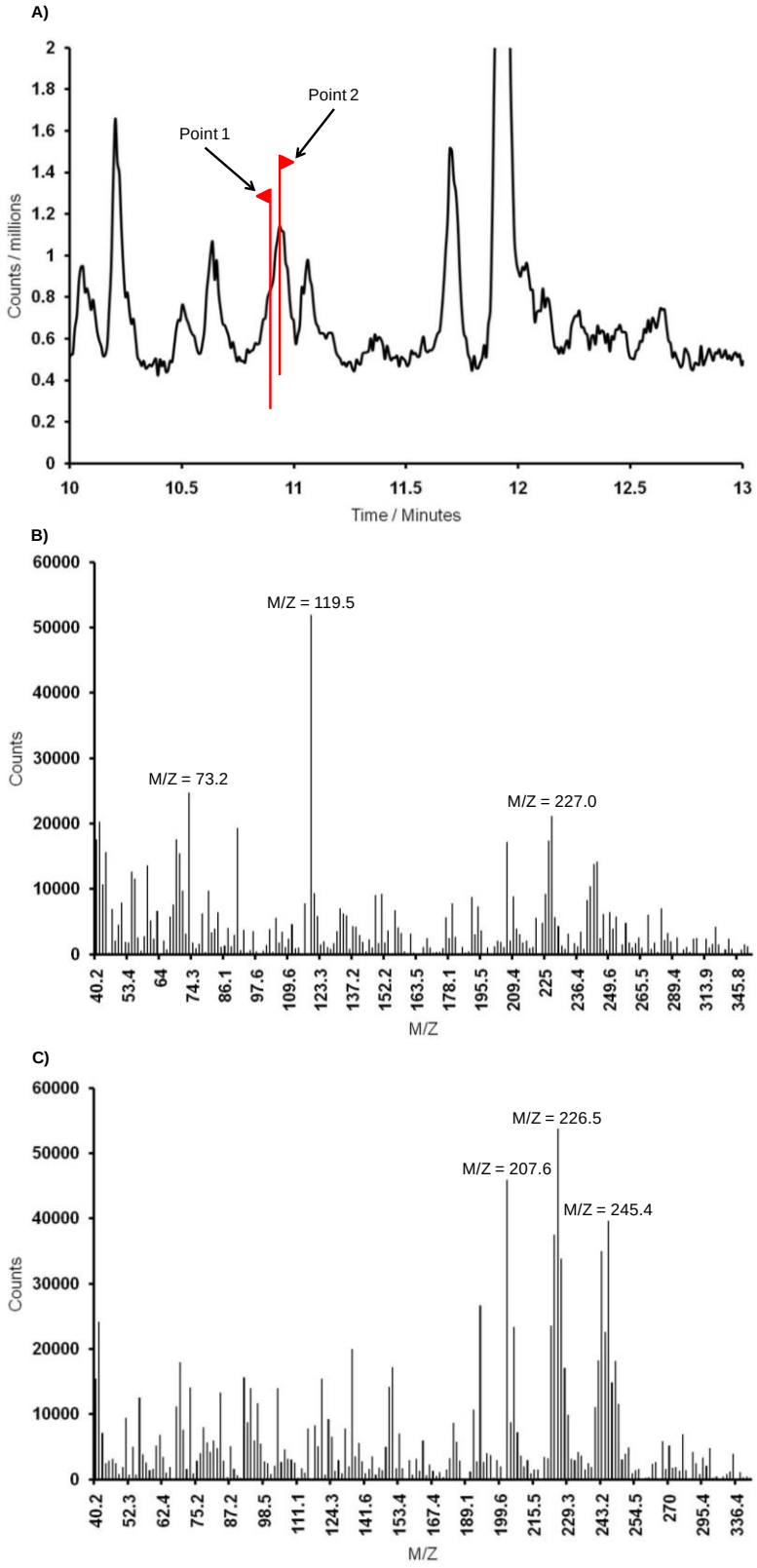


Figure 6-3 Example of how evaluation of the underlying mass spectrum was used to evaluate peak homogenietyis composed of a single or multiple components. A) shows a chromatogram with 2 points highlighted in a single peak. B) shows the mass spectrum at point 1 and C) shows the mass spectrum at point 2. While both spectrums B) and C) are from under the same chromatographic peak, the mass spectrum at both points is very different.

A list of components was created from the resolved peaks by selecting the underlying spectrum of each resolved peak at its most intense point and searching it against the NIST library. If this returned a plausible assignment it was added. A plausible assignment is one where the observed mass spectra under the peak matches the mass spectra of a suggested identification given by a library search. Peaks that were due to analytes not present in the NIST library were still added without any compound identification. Where mass ions that were indicative of a specific compound class were observed the component was labelled as belonging to that class. The six most intense mass ions, not due to common background ions, were recorded for each component. The common background ions excluded from the mass ions recorded are listed in Table 6-2 and Table 6-3.

Table 6-2 Observed ions, sources and possible compounds causing contamination peaks due to the analytical system.

Observed ions (m/z)	Source of contamination	Possible compound(s)
18, 28, 32, 44	Air leak	H_2O, N_2, O_2, CO_2
31, 51, 69, 100, 119, 131, 169, 181, 214, 219, 264, 376, 414, 426, 464, 502, 576, 614	Leak on tuning compound vial	Perfluorotributylamine (PFTBA)
69	Foreline pump oil vapour	Foreline pump oil

Table 6-3 Observed ions, sources and possible compounds causing artefact peaks due to the analytical system.

Observed ions (m/z)	Source of artefact	Possible compound(s)
73, 147, 207, 221, 281, 295, 355, 429	Septum and column bleed	Dimethylpolysiloxanes
14, 16, 44	Out-gassing of ferrules	N, O, CO_2
149	O-rings	Plasticiser

These component lists were then compared against each other to generate a single list of components found the collected breath samples and which samples they were found in.

Peaks that were found to be composed of multiple components were processed further.

6.2.5.2 Peak examination and qualitative separation of chromatographically unresolved analytes.

Unresolved peaks were manually analysed using Varian MS Workstation MS Data Review, Version 6.9 (Service Pack 1). This was done using a method based on the way unresolved peaks were determined in Figure 6-3. As previously shown in Figure 6-3, the mass ions under an unresolved peak change according to the concentration of the analytes eluting. By determining the correlations between ion intensities it is possible to extract individual component's mass spectra from co-eluting peaks. Figure 6-4 shows co-eluting compounds and the intensities of their associated fragment ions. The ions with m/z 77, 131, and 106 show a local maximum intensity at approximately 2.795 minutes while the ions with m/z 70, 55 and 83 show two local maximum intensities at approximately 2.77 and 2.86. This indicates that there are three co-eluting peaks under the TIC, shown in black in Figure 6-4, that can be separated based on the local maxima observed at different retention times. By removing all of the ions except those that show a local maxima at the identified retention time a "cleaned" mass spectrum for each separate component can be created and searched against a mass spectral database (NIST for example). This can be repeated for each analyte underneath an unresolved chromatographic peak until all analytes under the peak have been distinguished.

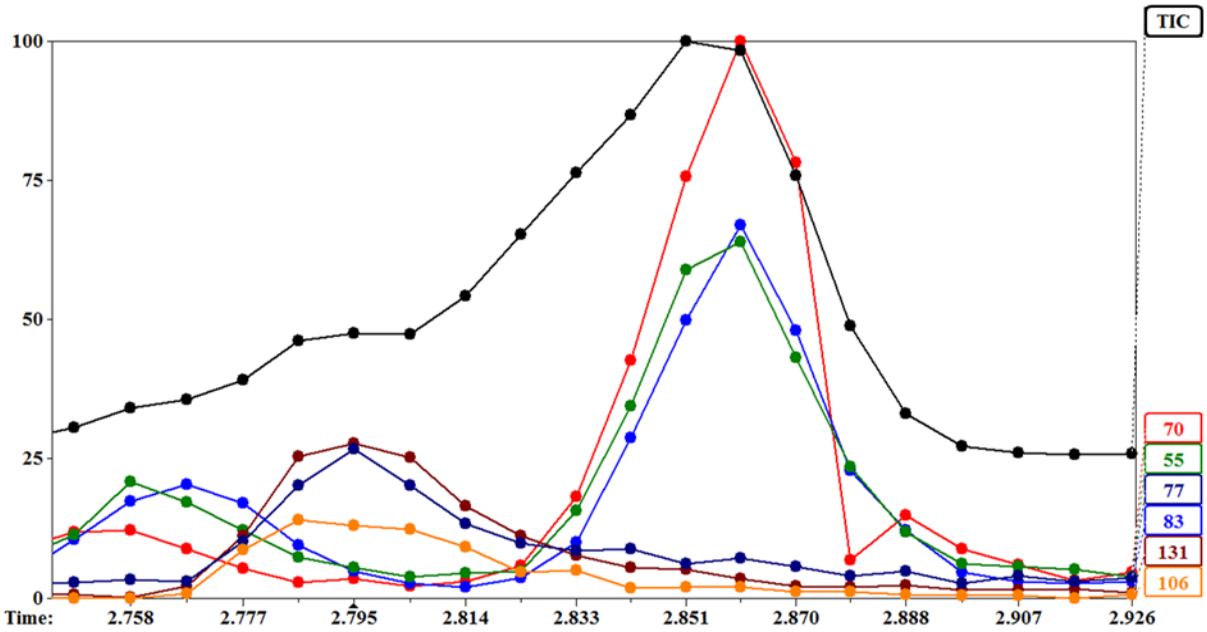


Figure 6-4 Example of how unresolved peaks are distinguished based on their underlying spectra. The peak at 2.795 minutes clearly contains different mass ions to the peak at 2.860 minutes. This example has been displayed using AMDIS for clarity.

6.2.5.3 Identification and detected volatile table generation

Once the chromatograms had been examined and the analytes separated under all of the unresolved chromatographic peaks, the cleaned mass spectrum of each analytes was searched for within the databases provided with the data processing package (Varian MS Workstation MS Data Review, Version 6.9 (Service Pack 1) NIST 05 Main Library),It is helpful to note that similar compounds have a similar mass spectra, and this challenge is exacerbated at low concentrations, or with high background interferences or both. Figure 6-5 shows the mass spectra of four methylated hydrocarbons, 2,2,3-trimethyl-pentane, 2,2,4-trimethyl-pentane, hexane and 11-(2,2-dimethylpropyl)-heneicosane which have similar mass spectra. Although 2,2,3-trimethyl-pentane, 2,2,4-trimethyl-pentane and hexane would be expected to have similar mass spectra, with some differences based on the different fragments observed. 11-(2,2-dimethylpropyl)-heneicosane is a much larger molecule, a larger molecule that has fragmented yielding the larger mass fragment-ions at a much lower intensity, if the background ions were particularly intense the larger fragments may not be seen and the compound may be incorrectly labelled as any of the other three.

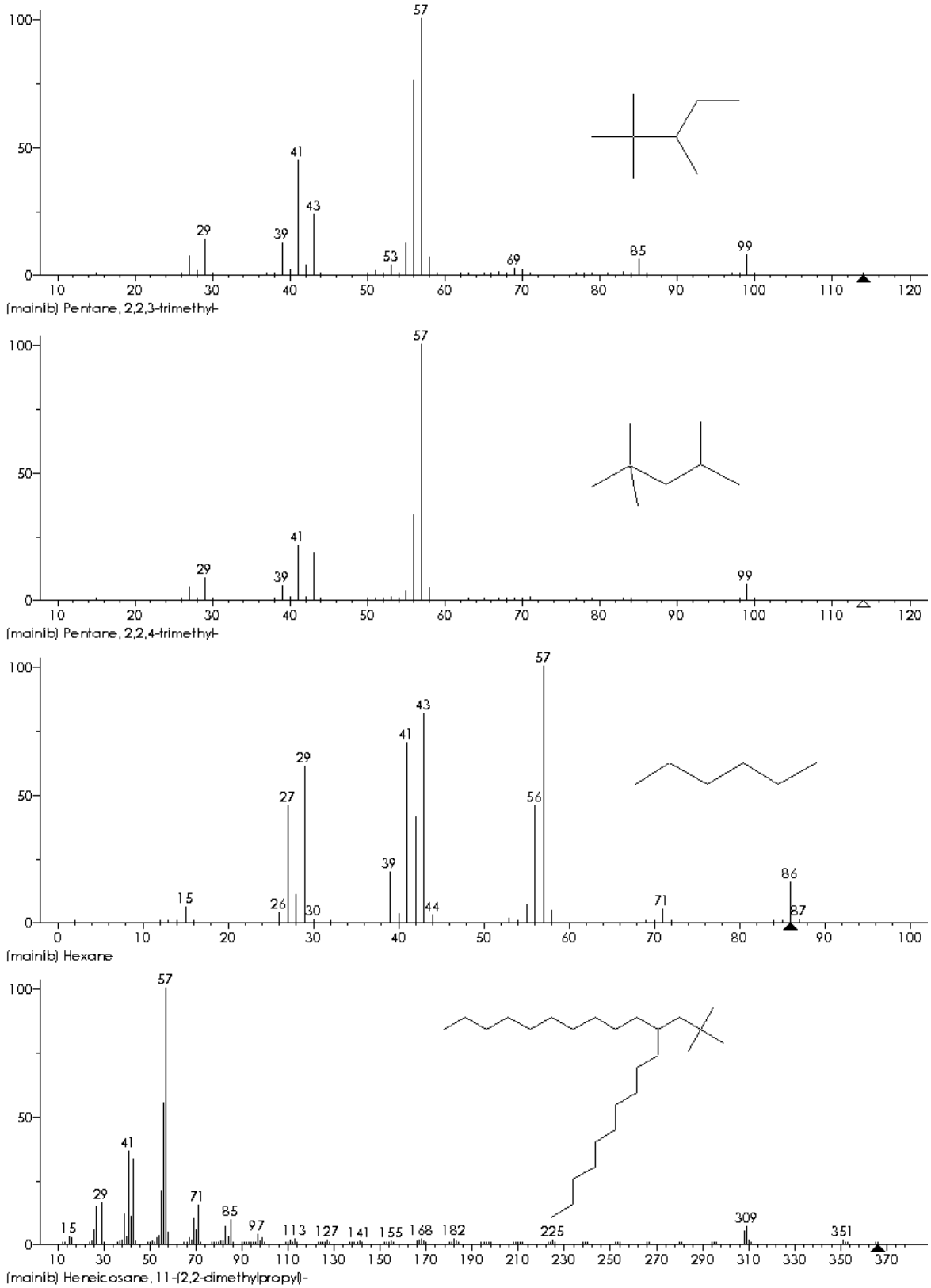


Figure 6-5 Four different compounds that produce similar mass spectra. 2,2,3- and 2,2,4-trimethyl-pentane are very similar, as would be expected. Hexane is the next most similar, showing that there is a possibility of confusion. The most intense fragments of 11-(2,2-dimethylpropyl)-heneicosane are very similar to those of both the trimethylpentanes and hexane. If a large background was observed it is possible that the larger fragments would be masked and may not appear.

So, it is important to remember that spectral libraries are based on mass spectra collected for well resolved standards at a strong concentration. This means that interference from background ions can influence any library identification and cause misidentification and so it is likely to be difficult to distinguish between isomers.

Consequently, each library hit was evaluated and the proposed assignments are attributed to those identities with high certainties, where several hits of similar molecules were observed that could not be distinguished a compound class was assigned instead of an identification.

A compound list was generated from each breath sample that containing all of the resolved analytes along with an identification if applicable and a list of the six most intense mass ions observed for the analyte along with their intensities. This was done in an effort to standardise the reporting of compounds found in breath so it is easier to cross-reference with other research. As two samples were taken per volunteer the lists from the two samples were combined to give a full list of compounds observed for a single volunteer. The two lists from each volunteer were evaluated and merged to give a single list.

Figure 6-6 shows a flow diagram which describes how the samples were processed after analysis. This was carried out for all of the samples collected from volunteers who were diagnosed with lung cancer.

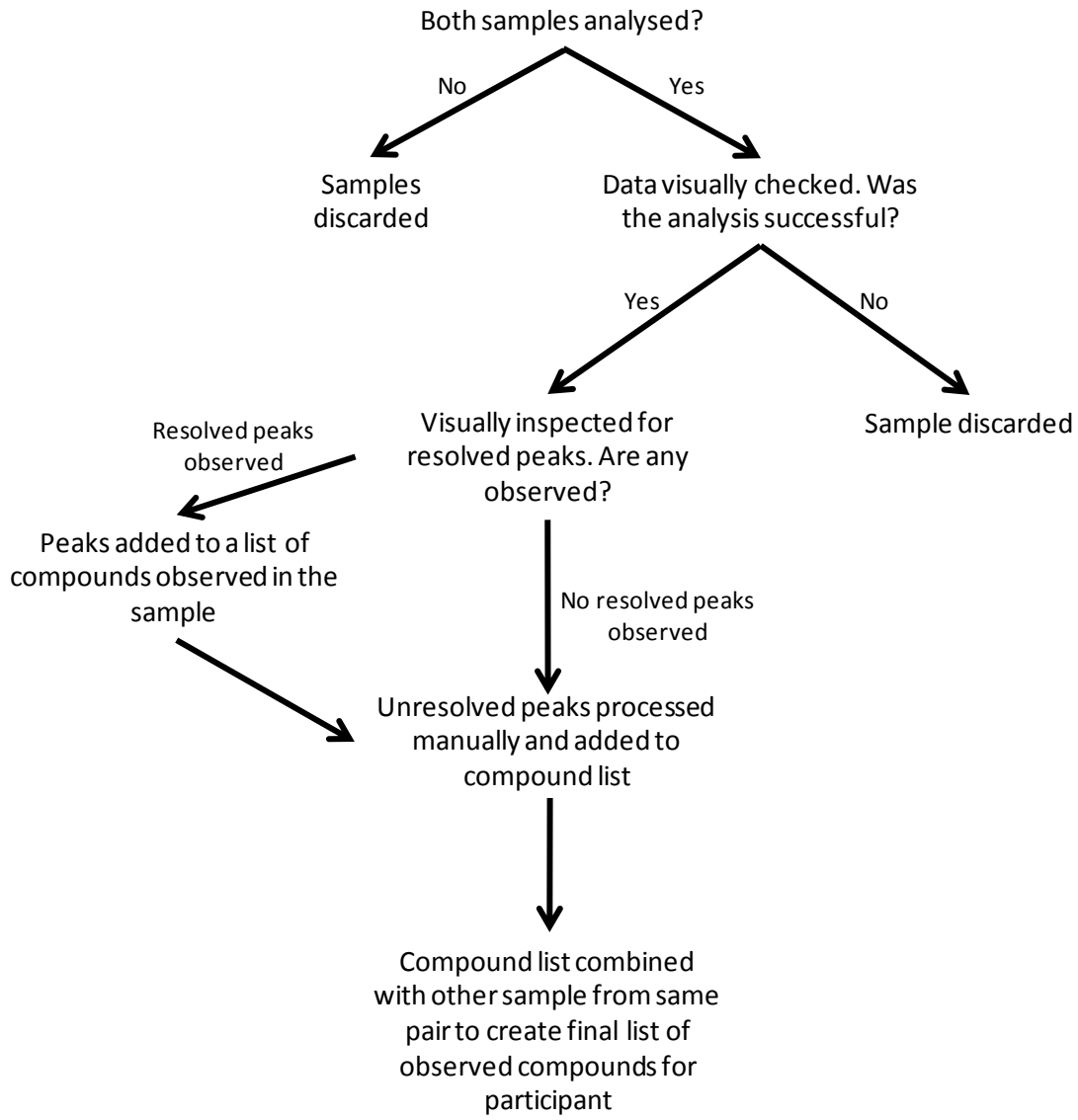


Figure 6-6 The flow diagram for how the data sets were processed.

6.3 Preliminary results from participants with lung cancer

For these preliminary results only the volunteers diagnosed with lung cancer were investigated. Volunteer CV-13 was excluded as only one sample was successfully analysed. Volunteers CV-05, CV-06, CV-10, CV-11, CV-12 and CV-13 were excluded as their diagnosis excluded lung cancer. These samples along with any others from volunteers whose diagnosis excludes lung cancer will form the control group when the study is completed. CV-07 and CV-09 could not be currently assigned to a group as a diagnosis of lung cancer has not been excluded or confirmed.

6.3.1 Preliminary data-processing

All of the samples were analysed on the instrumentation described in chapter 2 and visually inspected in MS Data Review. Large contamination/artefact peaks were observed in several samples. These samples were still included in the analysis but areas of the chromatogram covered by the contaminant/artefact peaks were excluded.

For each sample from a participant with a known malignancy a list of compounds was generated from the resolved and unresolved peaks using the methods described in Section 6.2.5.2. These lists were then combined to generate a list of compounds observed in the breath of lung cancer sufferers. Where no hit was obtained for the mass spectrum of the component this was either recorded as no hit or a tentative assignment of compound class was provided based on the most intense observed ions. To begin building a database of breath compounds and the associated participant diagnosis/information meta-data was recorded for each volunteer including their diagnosis and their smoking status at the time the sample was collected. Other information was collected as per the questions outlined in the participant questionnaire sheet.

6.3.2 Data processing

The aim of this initial inspection of the data generated from the pilot study is to describe the nature of the data, its' complexities and the best methods to address them. As well as differences between samples, there are other processes in the sample collection and analysis that may cause added complexity and these must be accounted for. The main challenges observed in the samples were:

- Differing intensities between sample pairs,
- Complex unresolved peaks,
- Foreline pump contamination and siloxane contamination and,
- Retention time shifting.

All samples were processed as described. To highlight the complexity of the samples the pairs have been plotted against each other and areas of complexity have been highlighted and discussed.

6.3.2.1 Differing intensities in sample pairs

From Figure 6-7 it can be seen that the intensity of peaks in the two samples collected from CV-08 are not consistent with each other. There are three reasons that this may be the case; the sampling was not as successful for the second sample, the analysis of the second sample using the instrumentation described in Chapter 3 was unsuccessful or the volatiles in the breath changed during collection of the second sample. While the volatiles observed in the breath are dynamic and may change from the collection of sample to sample it is unlikely they would undergo such a large change. The most likely case is that the adsorbent trap was improperly fitted to the mask. This could either cause a leak leading to less volatiles being drawn through the trap or it may have caused a block to the flow of air so that the pump was unable to draw the breath through the trap.

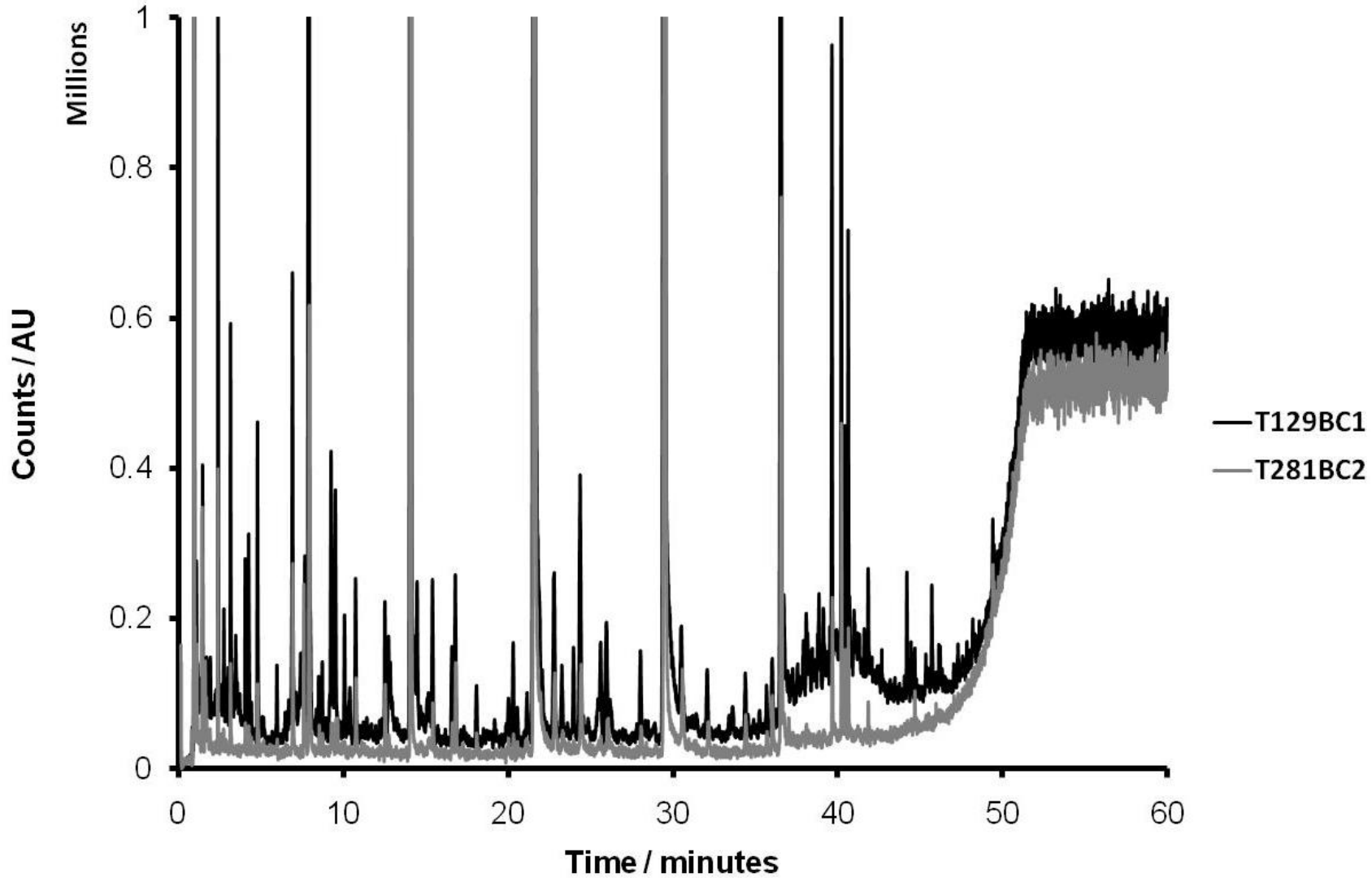


Figure 6-7 Showing the chromatograms produced by both samples collected from CV-08.

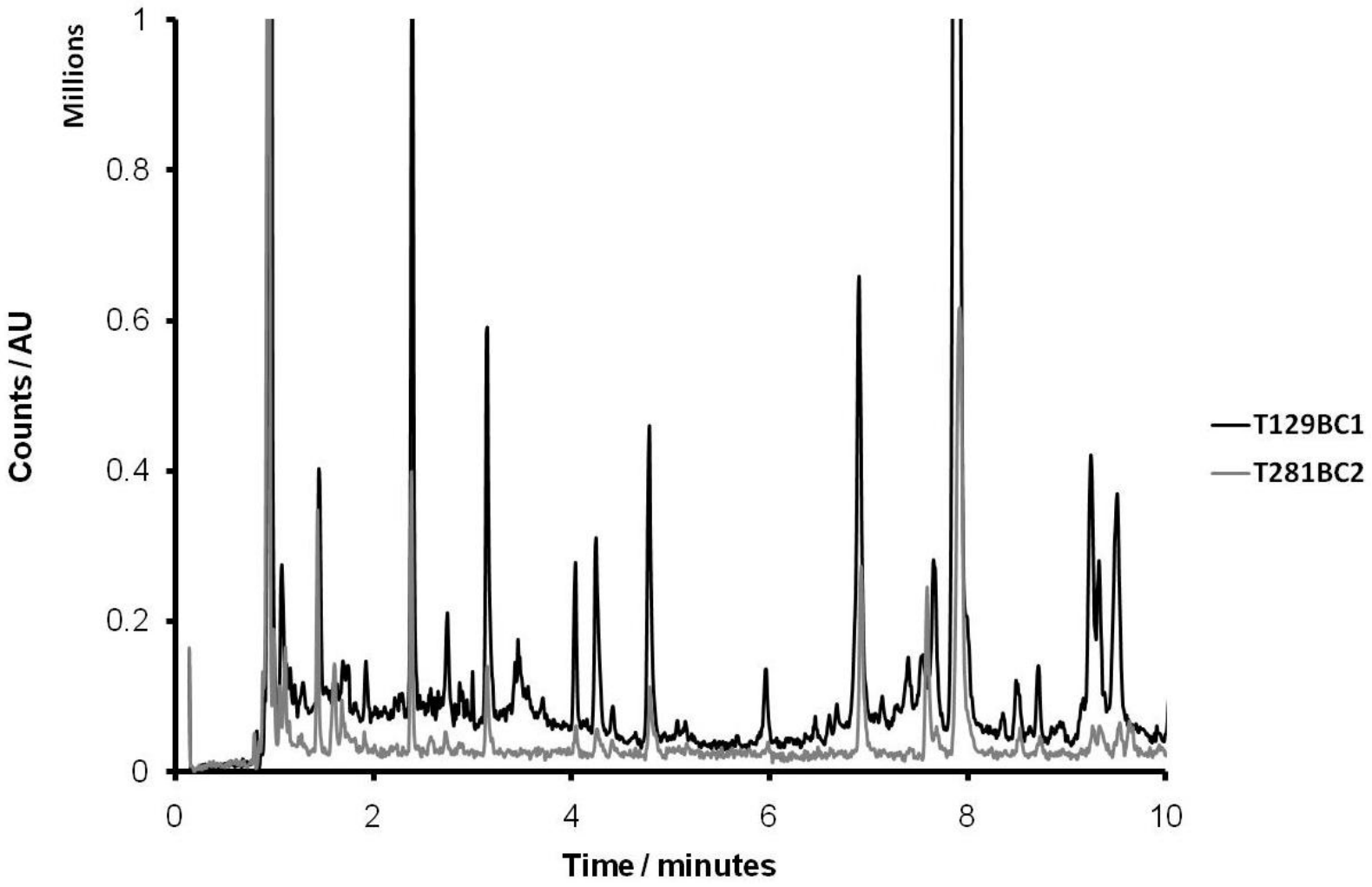


Figure 6-8 The first ten minutes of the analytical run for the two samples collected from volunteer CV-08.

Figure 6-8 shows the difference between the analysis of the first and the second samples collected from volunteer CV-08 for the first ten minutes of the analytical run. As can be seen the most intense peaks observed in the first sample (T129BC1) are repeated in the second sample (T281BC2). However, these are observed at approximately half to less than a quarter of the intensity of the same peaks in the first sample. These changes could be suggestive of either a problem during sampling or a problem during analysis. Figure 6-8 shows the problems that a loss of intensity can cause. Between six and ten minutes there are several peaks observed in the sample T129BC1 which are not observed in T281BC2. As the aim of this work is to search for unknown markers of disease, the apparent loss of peaks between two samples from the same participant taken consecutively is troubling.

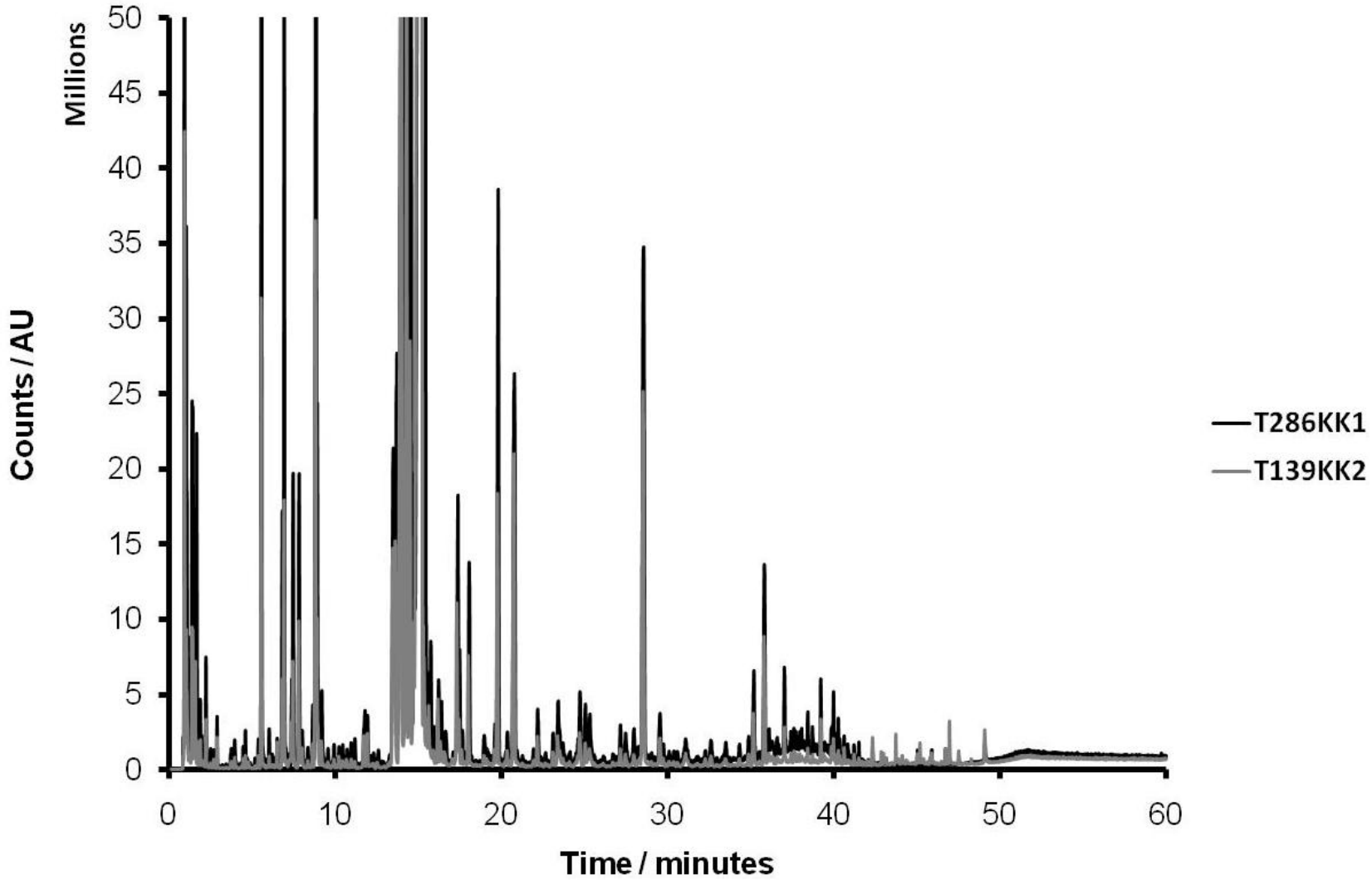


Figure 6-9 Showing the chromatograms produced by both samples collected from CV-17.

Figure 6-9 shows the chromatograms produced during the analysis of the two samples collected from CV-17. Similar to the analysis of the samples from CV-08, T139KK2 shows a lower overall TIC intensity when compared to T286KK1, although this effect seems less pronounced between these samples, with an average decrease in intensity of approximately half at its most pronounced. This effect is highlighted in Figure 6-7, Figure 6-8 and Figure 6-9.

Figure 6-10 shows both samples collected from CV-14. For the most of the chromatograms the intensities of the peaks are very similar, showing that the collection and analysis of the samples has been constant. This is more easily observed in Figure 6-11. It shows that the analytes are approximately the same intensity across both samples with some variation that can be accounted for in the slight differences in sampling and analysis that will occur. From approximately 42 minutes onwards a marked difference can be seen between the two samples. The peaks observed in this region are generally associated with foreline pump contamination. It appears that the contamination is much higher in sample T184JL1 compared to T282JL2. As the samples are sealed until analysis and kept at four degrees Celsius it is likely that this contamination occurred as a build up on the cold trap in the thermal desorption unit. This shows the importance of validating all samples manually before using automatic methods.

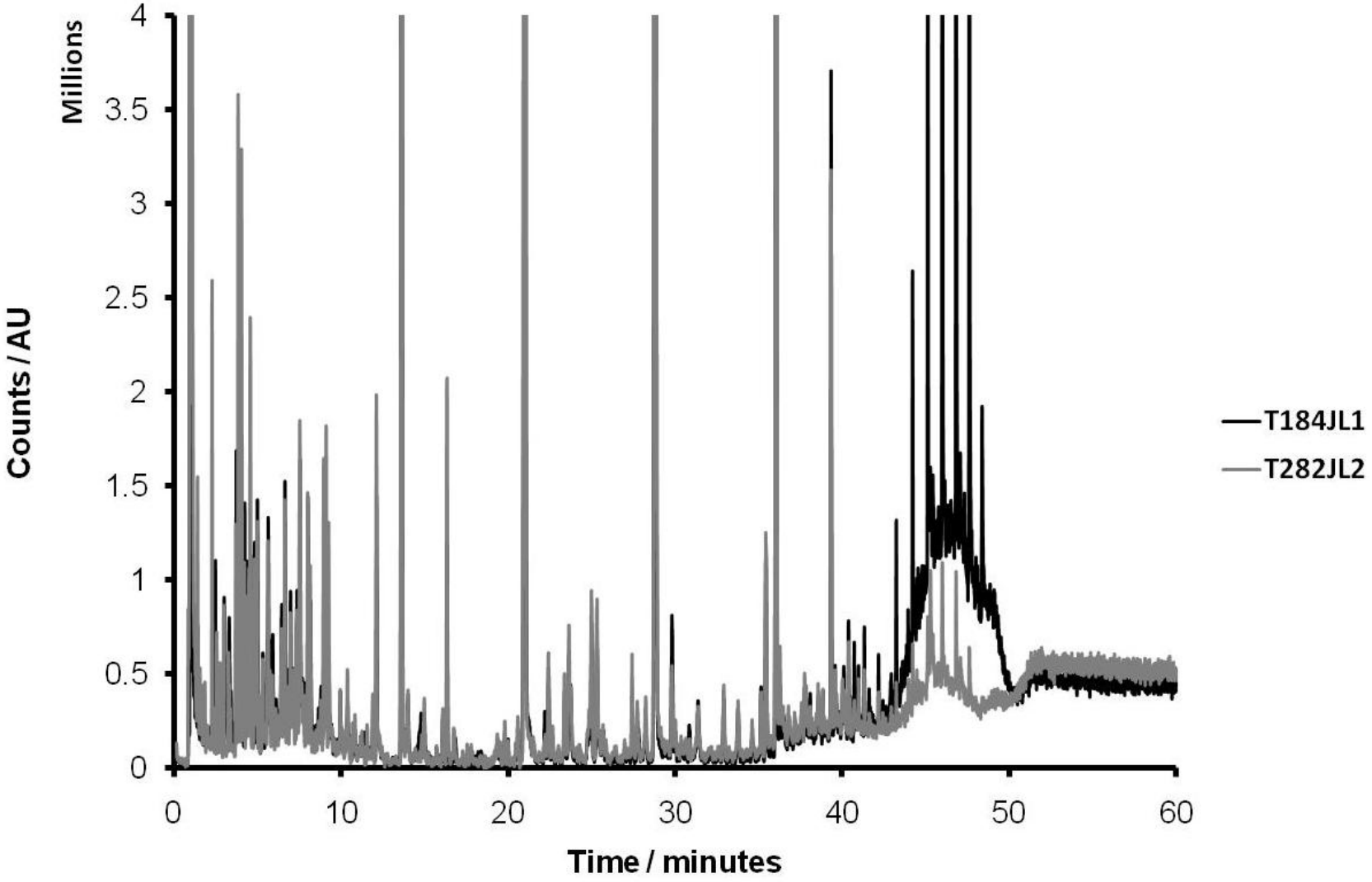


Figure 6-10 Showing the chromatograms produced by both samples collected from CV-14.

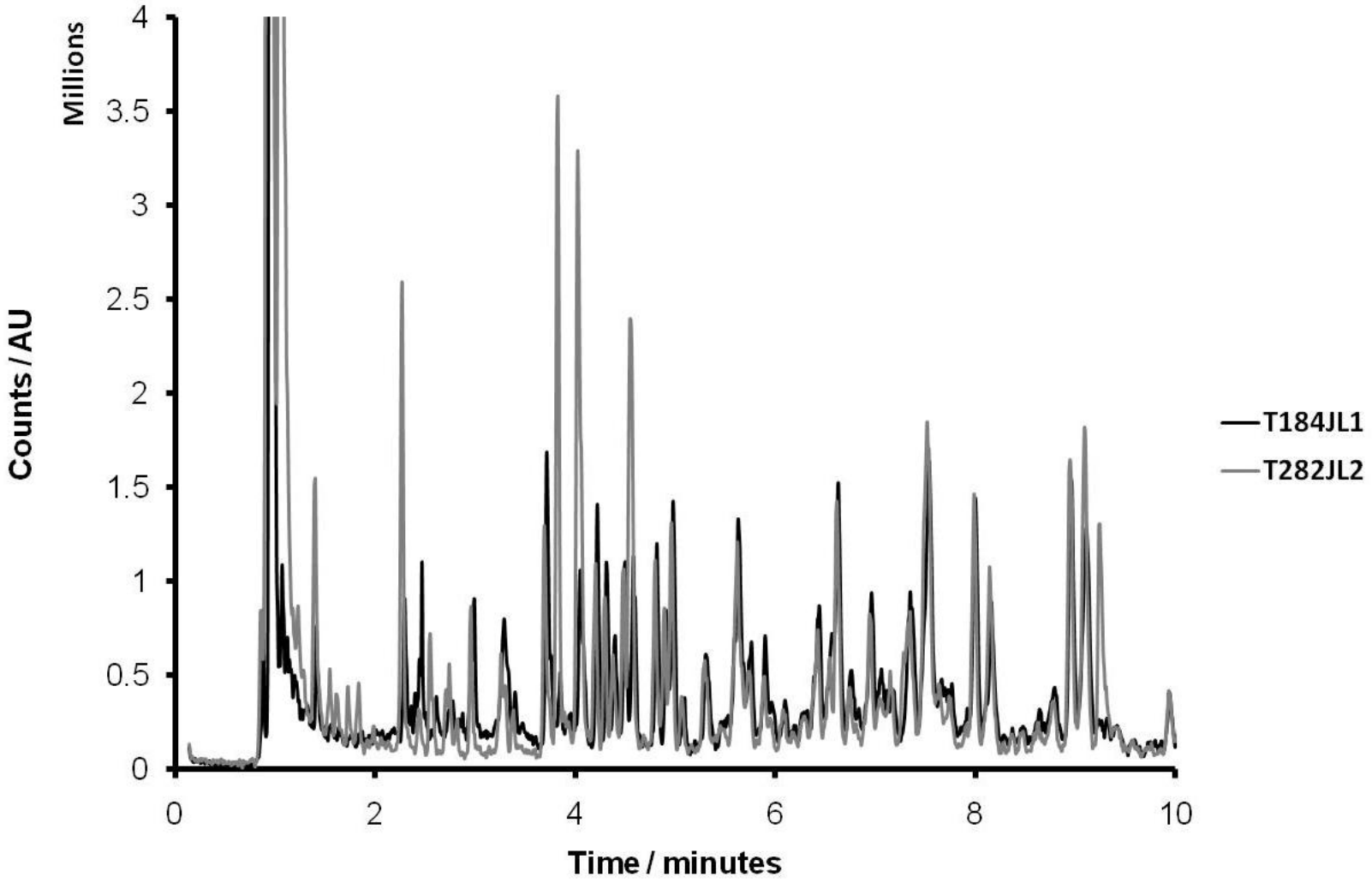


Figure 6-11 The first ten minutes of the analytical runs for the two samples collected from volunteer CV-14.

Figure 6-12 shows an increased intensity in the second sample collected from CV-16. This is opposite to the differences seen in Figure 6-7 and Figure 6-9. This indicates that it is more likely to be a sampler/instrumentation problem than a biological effect as the change is not consistent throughout the different sample sets. Again, the extent of the differences between the two samples can be more easily seen by examining the data more closely, as seen in Figure 6-13. This shows the extent of the differences between the two samples collected from the same participant and again highlights the need for manual inspection of the data, rather than a reliance on automated methods.

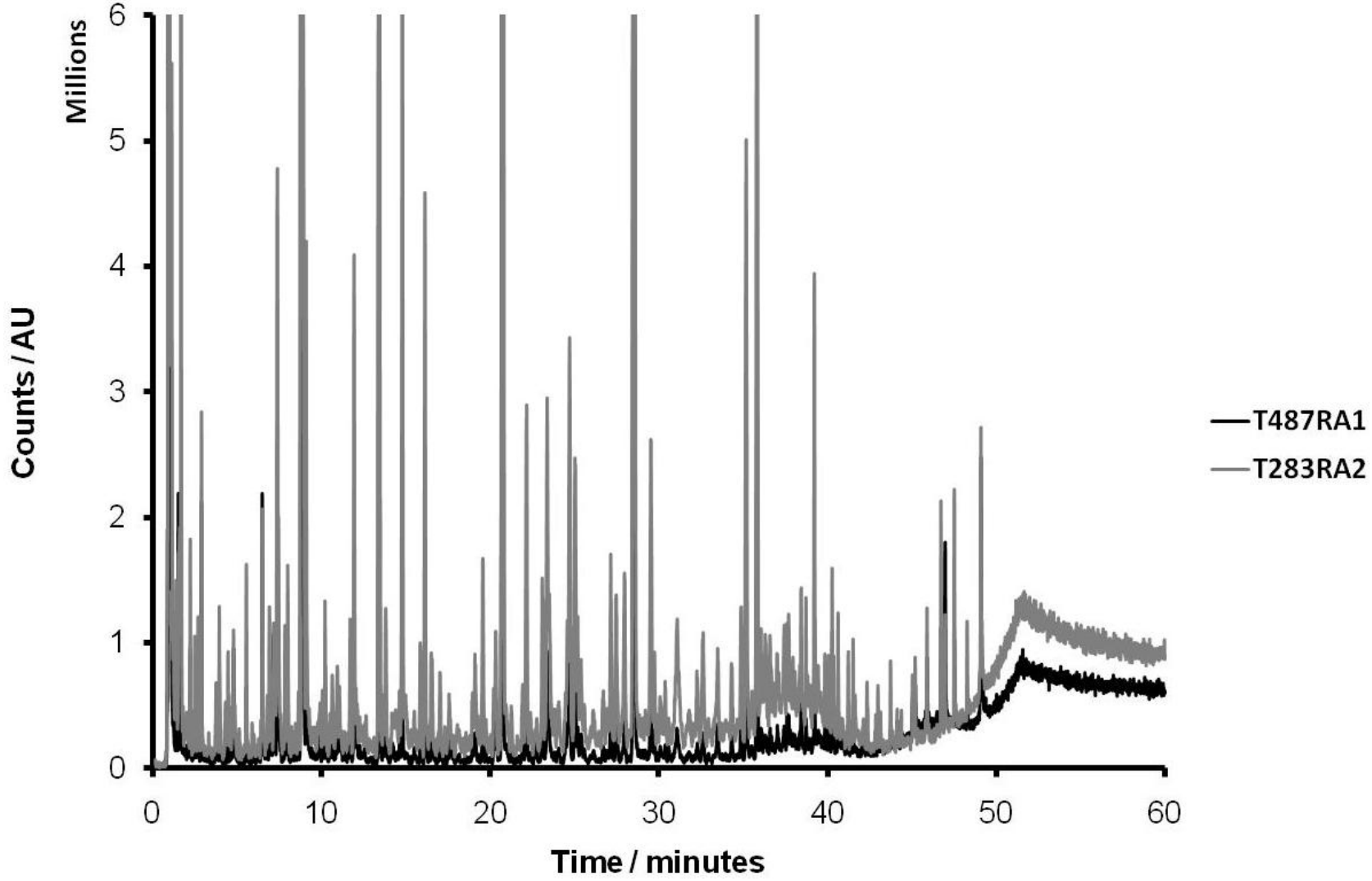


Figure 6-12 The chromatograms produced by both samples collected from CV-16.

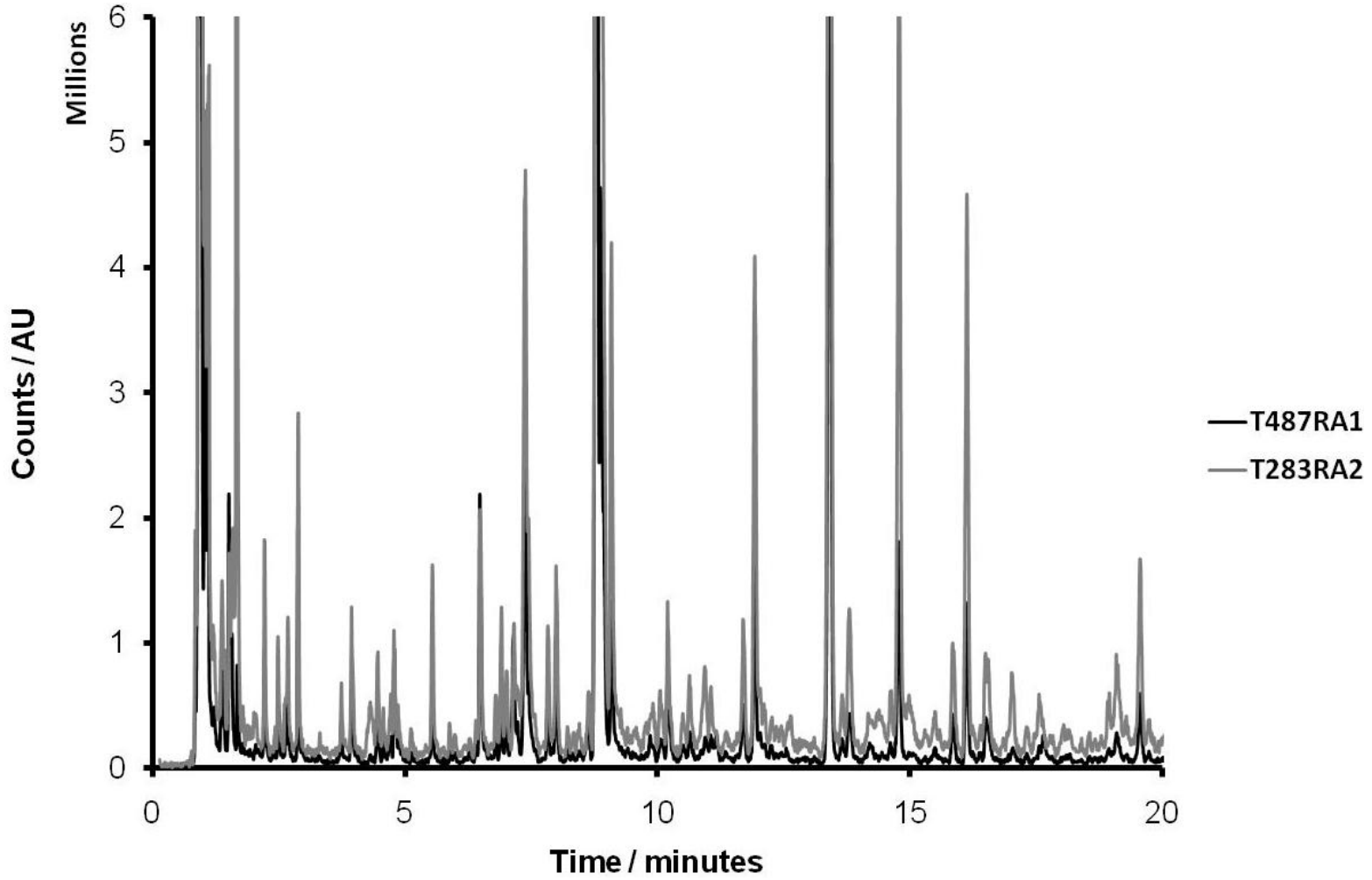


Figure 6-13 The first twenty minutes of the analytical run for the two samples collected from volunteer CV-16.

6.3.2.2 Complex unresolved peaks

Processing of complex unresolved peaks is the most challenging aspect of processing breath sample data. The method used to do this has been described in 6.2.5.2. Figure 6-14 shows the first four minutes of both samples collected from CV-17. This is a particularly complex region of the chromatogram, showing many unresolved peaks. This complex chromatography is seen in all of the sample pairs where the analysis was successful and made extraction of the single components difficult. To cope with this only components that could be satisfactorily determined based on the method previously described were considered.

Figure 6-14 is typical of the chromatography that must be dealt with when examining breath sample data. This must be dealt with carefully to make sure that no miss-identifications are made. With this in mind it may be better to use a retention index system and mass spectrum cataloguing rather than relying on identification of the separate compounds.

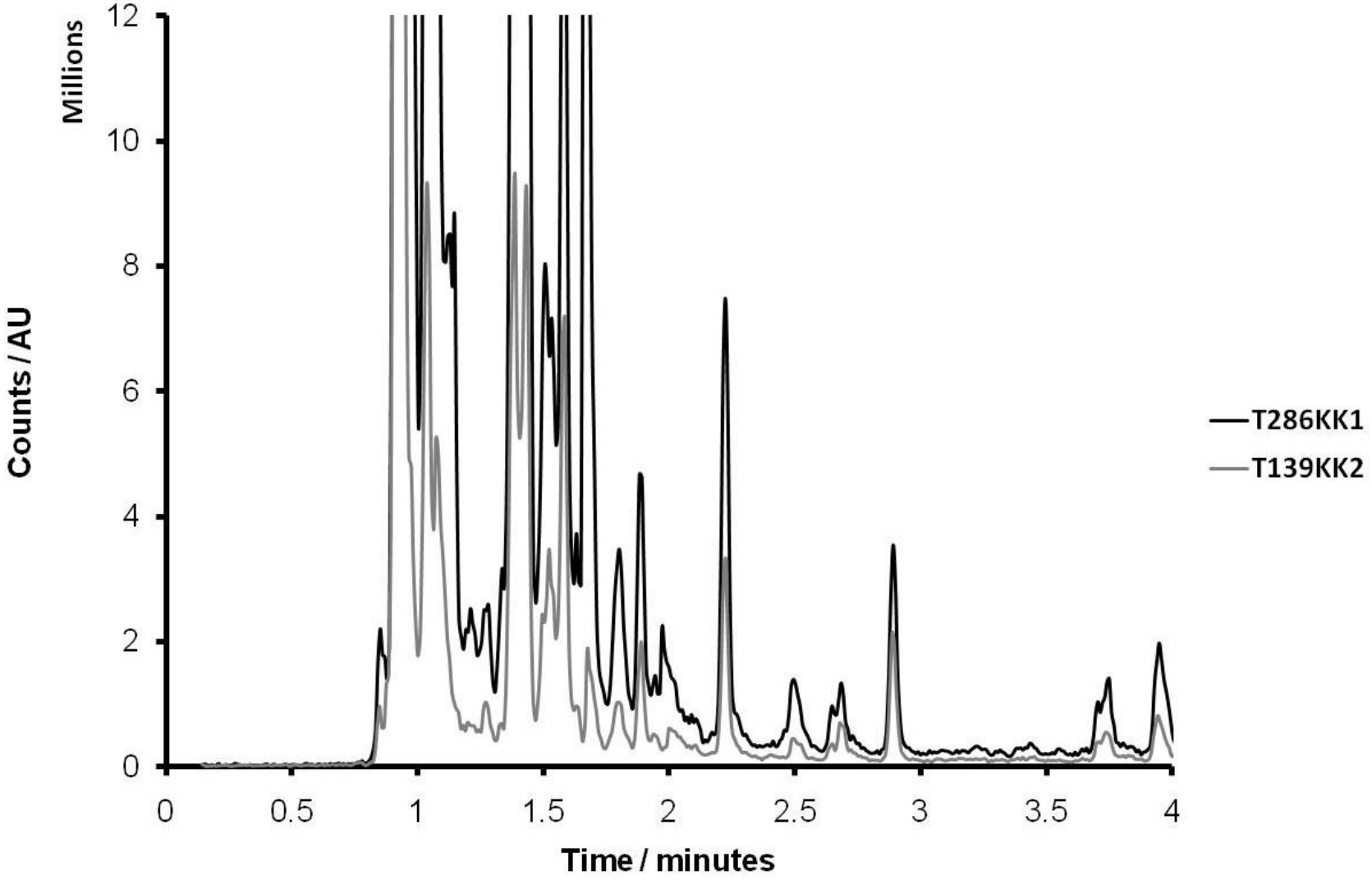


Figure 6-14 The first four minutes of the TICs produced during the analysis of the samples collected from CV-17..

6.3.2.3 Foreline pump contamination and siloxane artefacts and plasticisers

Figure 6-15, Figure 6-16, Figure 6-17 and Figure 6-18 show examples typical of foreline pump oil contamination. This can be a problem as some of the volatile components of the pump oil may be similar to those observed in the breath. The larger compounds observed due to foreline pump oil contamination are more easily distinguished due to their later retention times and higher molecular weights. Due to the possibility that compounds from the foreline pump oil may be mistaken for endogenous breath samples any compounds that elute close to the foreline pump oil contamination were ignored.

Differences in the intensity of the contamination peaks due to the foreline pump oil are observed for all of the sample pairs analysed. As this contamination occurs after sampling in the laboratory there are two possible ways the sample can be contaminated;

1. The adsorbent trap can become contaminated or,

2. The thermal desorber can become contaminated.

The adsorbent traps are kept sealed before analysis and when placed in the thermal desorber they are contained in a sealed system. This means that it is unlikely that the absorbent trap will be contaminated. The thermal desorber is open to the atmosphere during stand-by time. This is a prime opportunity for contamination of the instrument. A possible solution would be to seal the thermal desorber with a conditioned trap during stand-by time and monitor the contamination due to the foreline pump oil to see if it is significantly reduced.

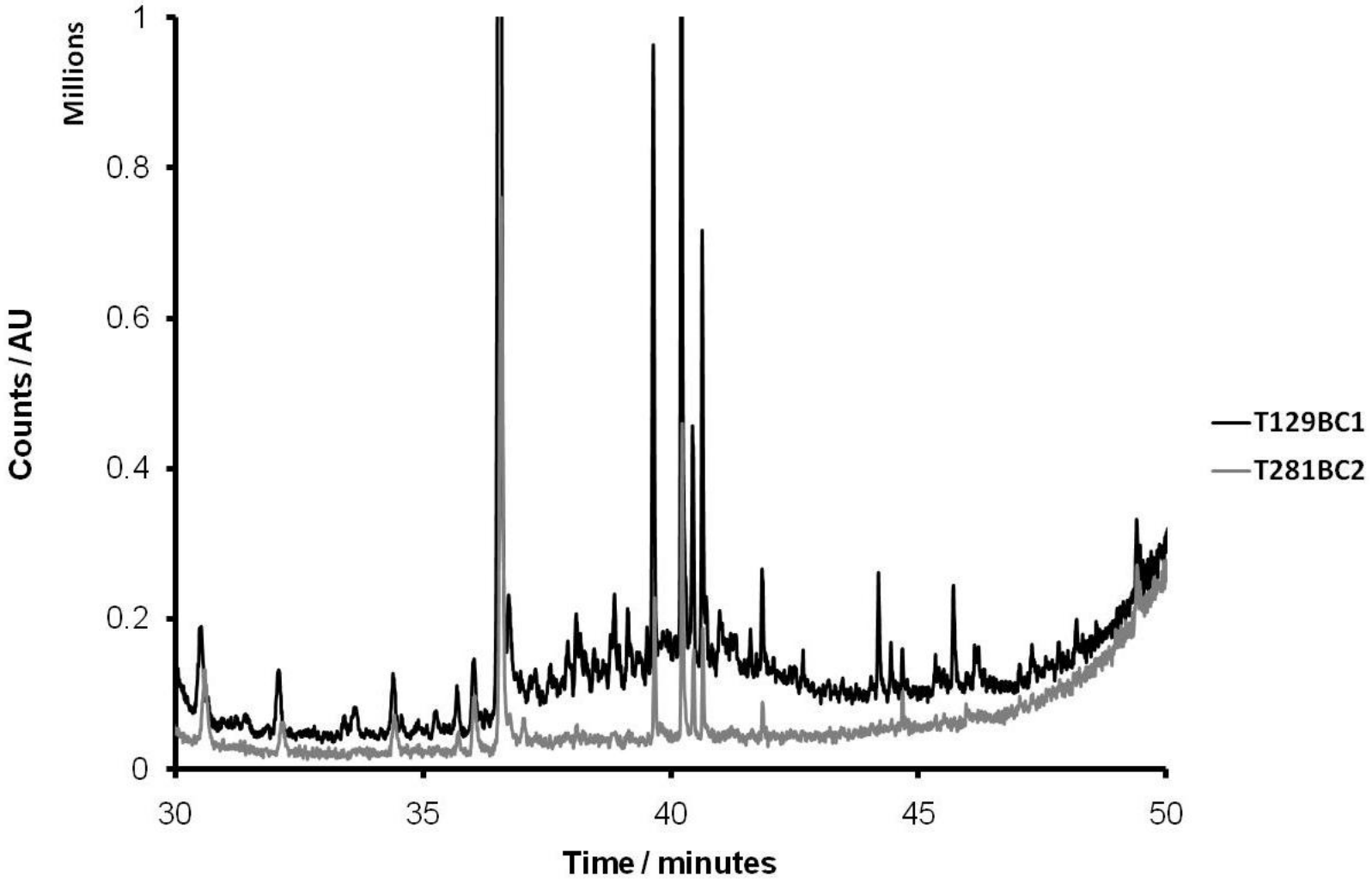


Figure 6-15 The twenty minute section of the analytical run from thirty minutes to fifty minutes.

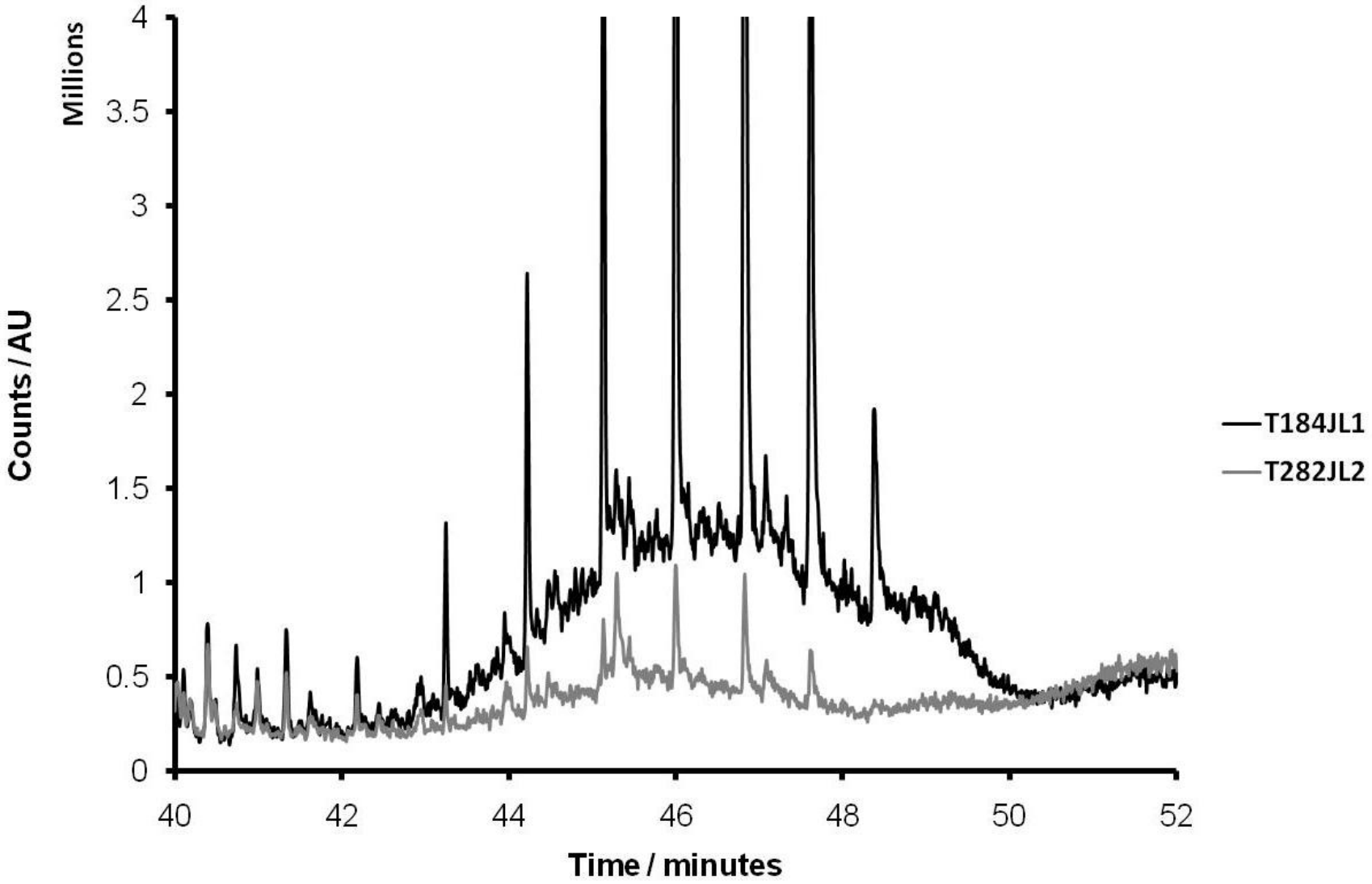


Figure 6-16 Shows the difference in contamination levels due to the foreline pump.

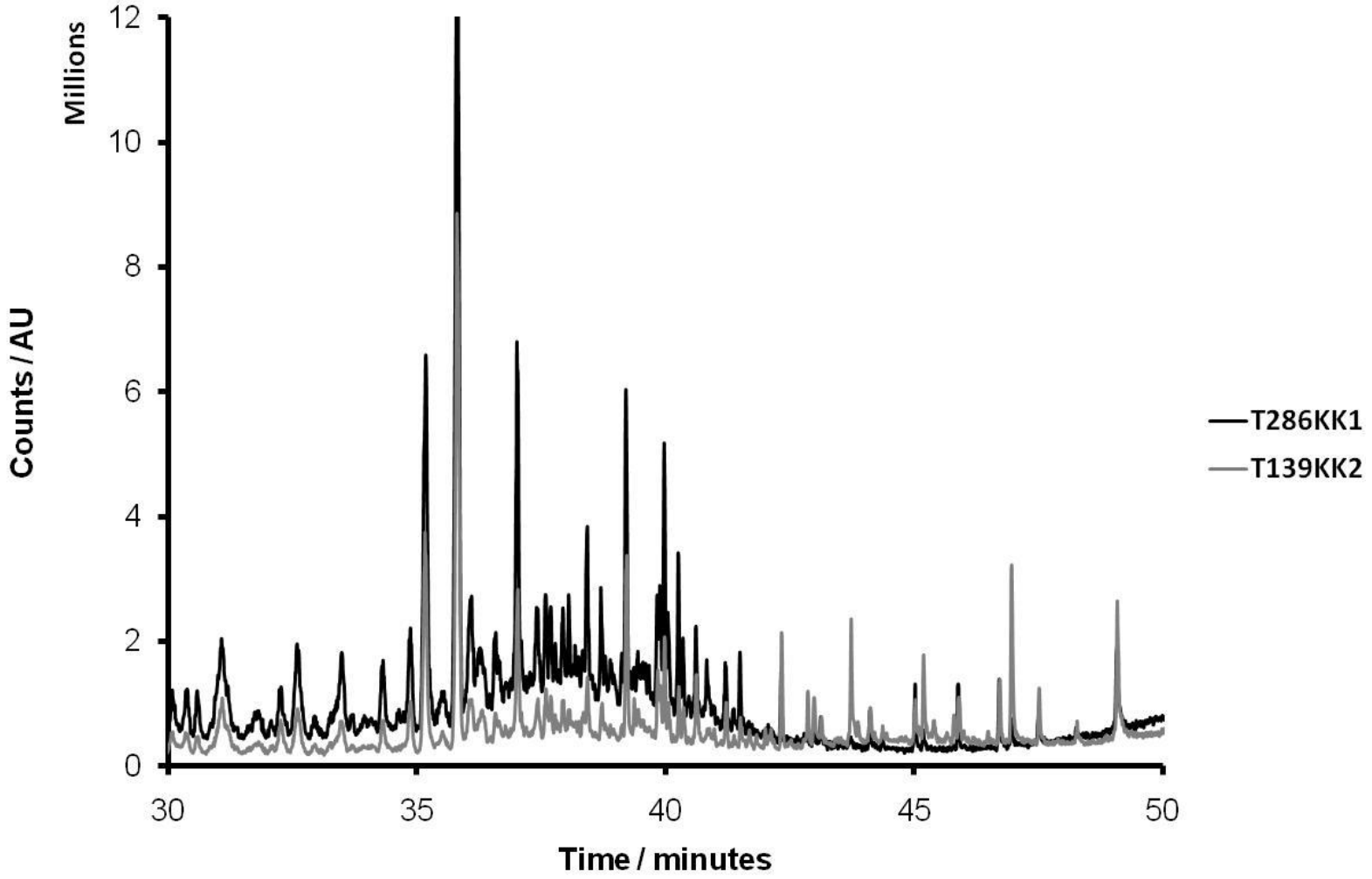


Figure 6-17 TICs produced for the samples collected from CV-17.

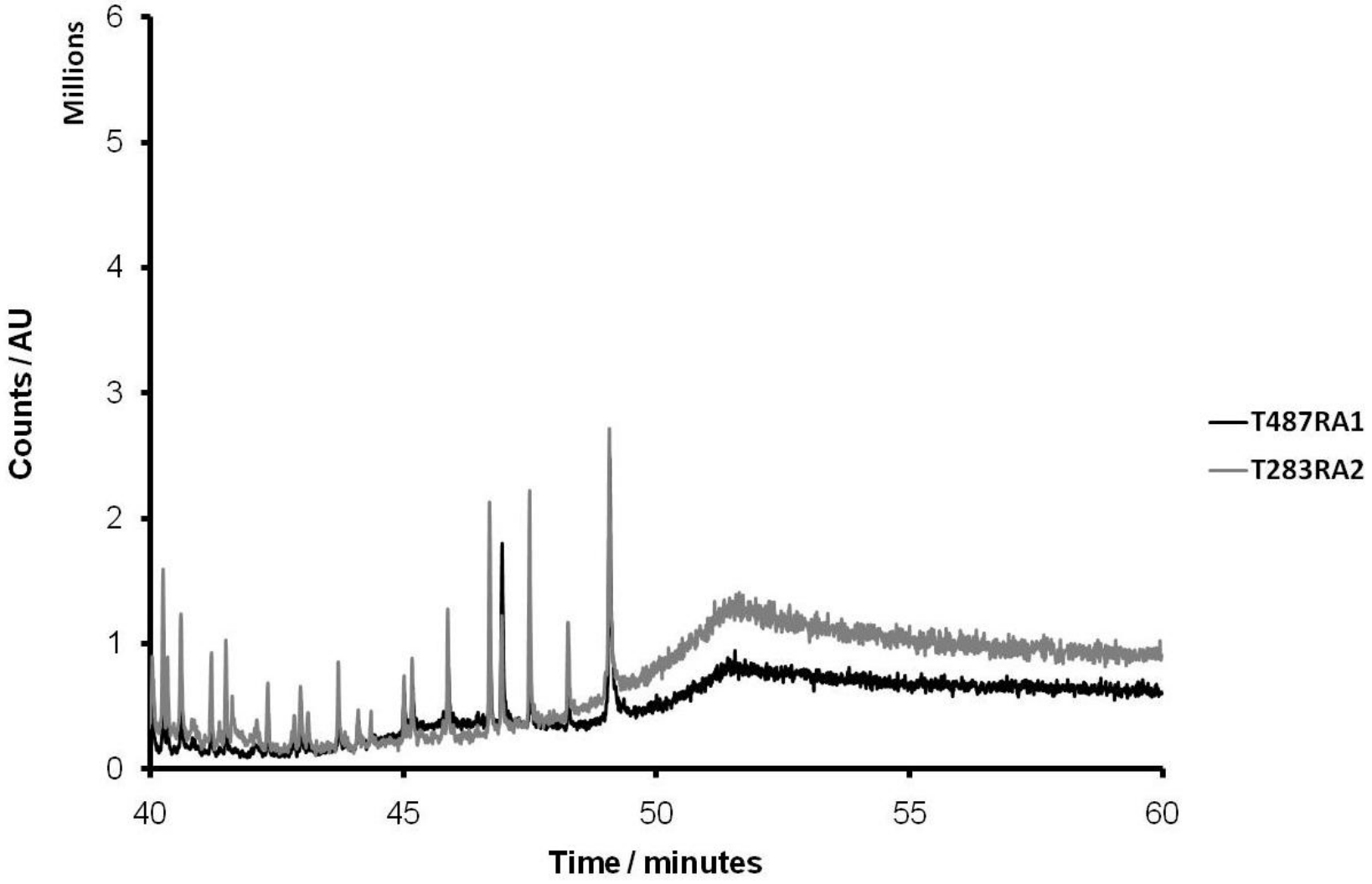


Figure 6-18 The final twenty minutes of the analytical runs for the two samples collected from volunteer CV-16.

Siloxane artefacts are seen in all of the samples analysed. It is likely that the high water content of the samples is a contributing factor to the large siloxane peaks observed in the samples. Table 6-3 Observed ions, sources and possible compounds causing artefact peaks due to the analytical system. lists several mass ions that are associated with siloxanes, m/z 207 being the most common. With experience it can be relatively easy to spot siloxane artefact peaks and these peaks may even be useful to help align data sets. Small resolved siloxane peaks do not present a problem and may even be beneficial but large, unresolved siloxane peaks can be a problem. Figure 6-19 is an example of a large siloxane peak observed in the samples from the participant CV-17. This peak may be co-eluting with other peaks that are masked due to its' size and these peaks are dealt with in a similar way to foreline pump oil contamination peaks; unless and accurate identification can be given to a compound that co-elutes with a siloxane peak, the peak is ignored.

Plasticiser peaks are caused by volatile compound bleed from septa, o-rings and other plastic parts within the instrument. As with siloxanes there are distinct mass ions associated with them, particularly m/z 149, and they are usually easy to spot with experience. These are treated in the same way as siloxane artefact peaks.

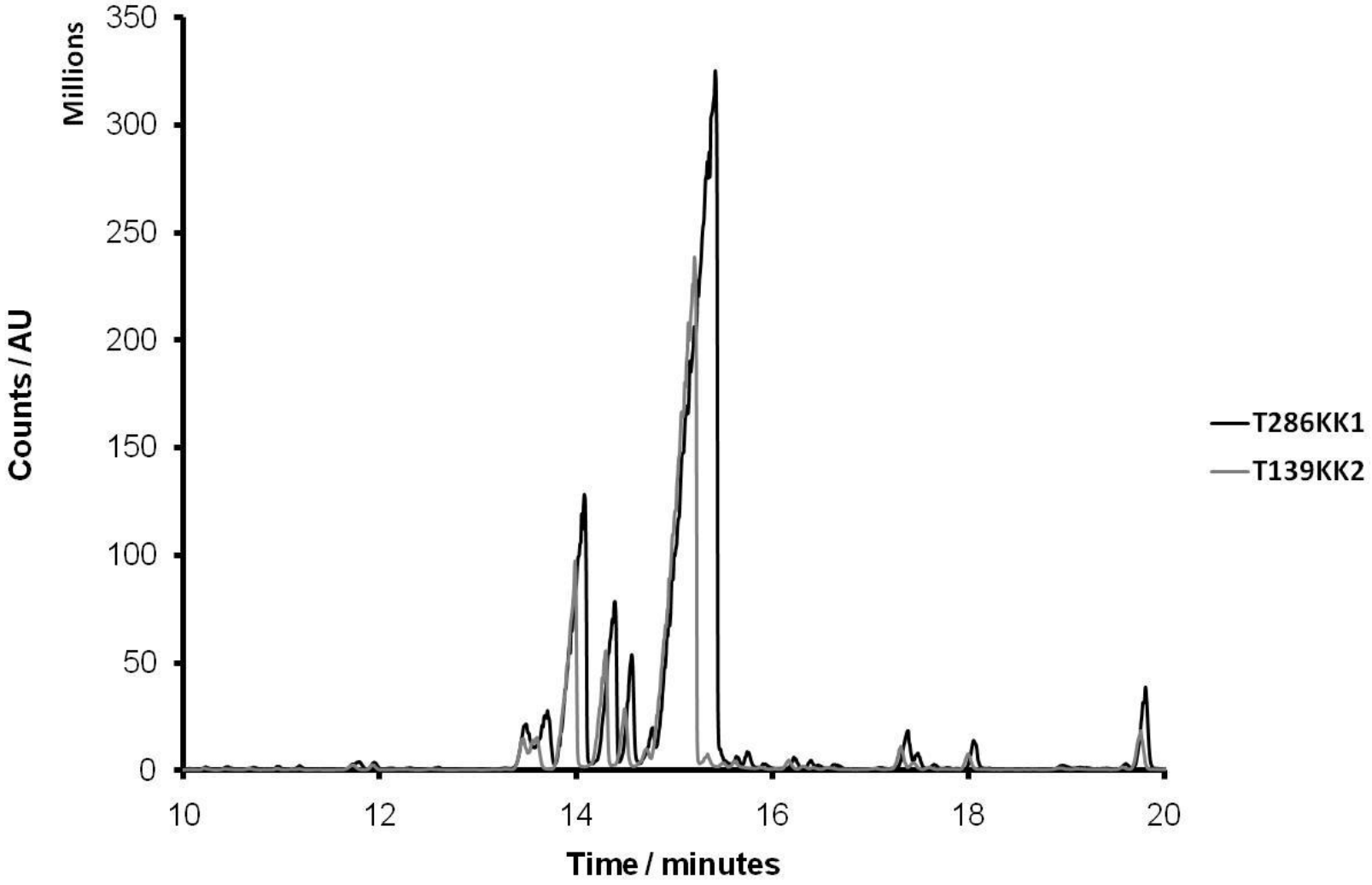


Figure 6-19 TICs from the two samples collected from CV-17.

6.3.2.4 Retention time shifting

Retention time drifting caused problems with deciding if peaks were reproduced across all or some of the samples. Retention time drifting is when the same analyte appears at slightly different retention times in different samples. Figure 6-19 shows an artefact siloxane peak from a single sample from each volunteer. As previously seen in Figure 6-7, Figure 6-9, Figure 6-10 and Figure 6-12, retention times were reproducible over the pairs of samples. However, it can be seen in Figure 6-20 that retention times are not reproducible across the analysis of every sample. This means that compounds which could not be identified while creating the compound list for each set of samples may appear in several of the samples but cannot be positively identified as such due to retention time drift.

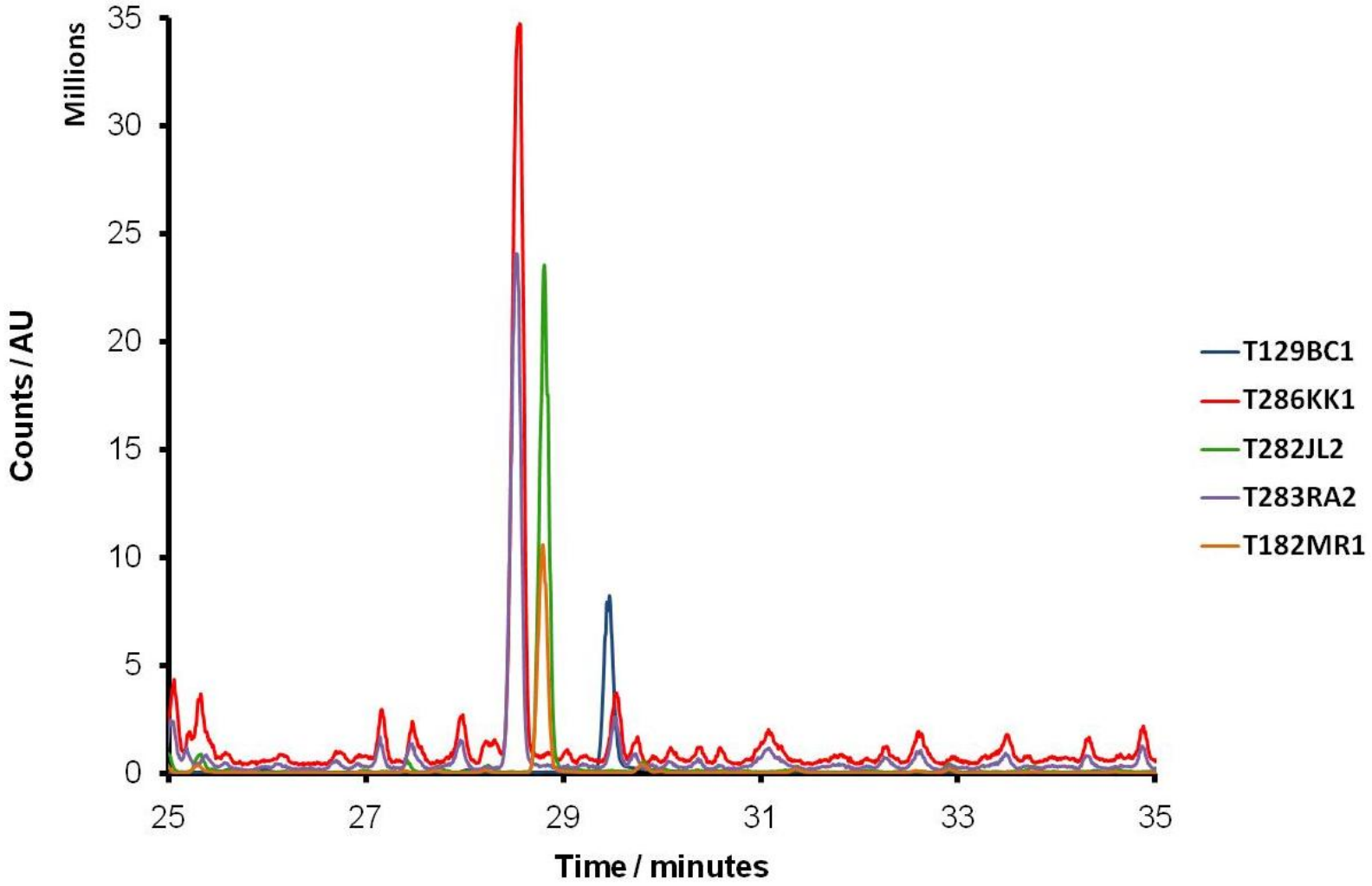


Figure 6-20 The chromatograms for one of the two samples from each pair of samples for participants diagnosed with lung cancer, including the only sample analysed for CV-13. The main peak shown is an artefact siloxane peak common to all of the samples to show the drift observed for the same compound in different samples analysed at different times.

Sample T283RA2 and T286KK1 show a very similar retention time for the observed peak and were collected on the same day and analysed in the same batch. Samples T282JL2 and T182MR1 also show a similar retention time and were within two weeks of each other. Samples T283RA2 and T286KK1 were collected one month after T282JL2 and show a small variation in retention time with the siloxane peak being less well retained in the later samples. Sample T129BC1 shows a larger retention time difference for the siloxane peak, with it being more retained. This shows a trend for a decrease in retention time over the course of the analysis, possibly due to deterioration of the column.

For complex samples where many analytes are similar and unresolved retention time drift can be a large problem. As previously discussed in (CROSS REF SIMILAR MASS SPEC HEADING) many compounds have a similar mass spectrum, particularly at low concentrations and in samples with high background noise. As similar compounds will elute from the column at similar retention times, changes to this retention time could cause compounds to be confused for each other when comparing across samples.

6.3.2.5 Addressing the complexities and identifying compounds found in the breath of volunteers diagnosed with lung cancer.

All of the samples were analysed taking into account the complexities of the data outlined in the previous sections. For each sample a table of compounds was generated. This means two tables are generated for each participant, one for each sample. These two tables are then combined to create a single table for each participant.

Once a table of compounds had been generated for each participant (Appendix VI Table 1 – Table 4) these were combined to create a table of all the compounds found in the breath of participants diagnosed with lung cancer. Where possible compounds that were observed in more than one participants samples were combined, such as benzene observed at approximately one point three five nine minutes, o-xylene and p-xylene observed at three point eight five three and four point zero three seven and the “no hits” compound observed in all of the samples at four point four five seven. Where peaks were not obviously the same across the samples they were included separately in the list.

The final list of compounds observed in the breath of participants with lung cancer can be seen in Appendix VI Table 5. The intensities of the observed mass ions are not included in this table for brevity.

6.4 Discussion

6.4.1 Compounds found in the breath of volunteers with lung cancer

Appendix VI Table 5 contains a list of all of the compounds found in the breath of volunteers diagnosed with lung cancer. This will form the basis for a database of compounds found in human breath. The compound list contains a large number of compounds that could not be identified using the search tool for the NIST database. There are several reasons why this may be;

- The compound is not listed in the NIST database and so no reasonable match can be found.
- High background interference may be masking important mass ions and so hinder identification.
- Changes in ionisation may not provide necessary mass spectrum for identification.
- A large number of hits were given by the search with no obvious candidate.

To deal with these possibilities several methods were used. When no identification is possible through the NIST program a tentative assignment of the compound class, such as hydrocarbon, benzene derivative et cetera, based on the most intense mass ions can be made. If this is not possible then the compound was still included but with no identification.

When a large number of related hits were given, such as for the methylated hydrocarbons identified, only the compound class was given as it was felt that accurate identification of the compounds was not possible. Contamination/artefact peaks

There are several different stages where contamination of the samples is possible and it is also necessary to note the difference between contamination and artefact peaks. A contamination peak is a peak due to an exogenous compound entering the sample at some point in the process. This can occur at any point from when the sample tube is cleaned to when it is analysed. The compound class of the contaminant may make it easy to exclude, such as compounds containing chlorine in a biological sample. While hydrochloric acid is present in the stomach, this is a reasonable assumption as the reaction kinetics of chlorination of phenol, a compound likely to be chlorinated are unfavourable at pH 1-3 [169] as observed in the stomach. However, certain contaminants, particularly hydrocarbons may be impossible to distinguish from the sample, particularly in a dynamic sample set such as breath samples. The best way to reduce the possibility of contamination is to minimise the points in the sampling and analysis protocols at which contamination is possible and to minimise the chance of contamination at these points. This is done by developing your sampling technique in a way to eliminate exogenous compounds, such as using a filtered medical air stream to feed a non-vented full face mask as laid out in the protocol for using the adaptive breath sampler [104]. Other methods include analysing samples in a timely fashion, in this study forty eight hours after collection was set as the maximum time a sample could be stored for prior to analysis, and using suitable transport and storage methods, such as maintaining the samples at a temperature equal to or below four degrees centigrade. Contamination sources in the analytical system include poor quality carrier gas, dirty carrier gas tubing, foreline pump oil vapour and leaks on the tuning compound vial or column connections. System maintenance and monitoring is the best way of controlling system contaminants. A list of common contaminants arising from the analytical system is shown in Table 6-2.

Artefact peaks are peaks due to internal factors; in the case of this work they include bleed from degradable parts of the analytical system and, in particular, the degradation of Tenax to benzene. Some artefact peaks can be easily spotted based on the ions observed in their mass spectrum. A list of common artefact ions is included in Table 6-3. Artefact peaks are typically caused by column and septum bleed, out-gassing of ferrules and plastic parts such as o-rings. While factors such as the amount of water in the system and the highest

temperature that the column is raised to during the run determine the size of the artefact peaks, they tend to be present in all runs so are more easily distinguished compared to contamination peaks.

As the pressure on the chromatography columns is higher than usual due to fact that both analytical columns need to be driven by a single flow it is possible that this is causing an increase in both contamination and artefact peaks observed during the analysis of the samples. There are two ways this can be dealt with:

1. Post-processing to remove the peak

This method would allow the system to operate as it currently is, however, with large contamination peaks observed it may cause analytes to be lost if they were to co-elute with these peaks.

2. Split the samples between two analytical systems, one GC-MS and one GC-DMS.

This method would solve the problem of having to use a very high pressure to drive two analytical columns simultaneously, but would cause a loss in sensitivity. This method is probably the best compromise and should be explored.

6.4.2 Normalisation of data sets

One of the main problems with breath data is the sample to sample variability. When purely looking to identify compounds and possible differences between two groups this may not present a large problem but if this data is to be used for any statistical analysis this needs to be addressed. Normalising the final data sets is imperative and identifying what to normalise against will be important. The options available are:

- Normalise against the largest peak in all the data sets. This may cause some smaller peaks to be lost, particularly if the average intensity of some of the data sets is significantly lower than the most intense, as has been shown in this work.

- Normalise against the same peak in every data set. This would restrict normalisation to within each data set and so smaller peaks will not be affected by normalising against an external peak, however, identifying a peak which is consistent throughout all of the data sets is difficult.

- Normalise against the largest peak in the data set. This would be the most likely option as it restricts normalisation to within each data set while varying which peak is used to normalise against keeping it constant.

Of these options the most plausible would be to normalise against the largest peak within each data set. This would remove the influence of large intensity differences between data sets caused by the instrumentation or sampling equipment and would also mean that there was no need to identify a consistent analytical peak.

6.4.3 Manual versus automated deconvolution methods

There are both manual and automated deconvolution methods that can be used for deconvoluting complex GC-MS mass spectrometry data. There are advantages and disadvantages to both manual and automated methods. Manual methods allow the person processing the data to decide if a peak is composed of one or more components and which ions belong to which component. This can be a long and difficult task depending on the data set being processed and there is the chance that bias can be added to the separation of the peaks. However, as the researcher is able to use their knowledge to decide which ions belong to which component of the peak, it can be a much more accurate method than automated processing.

The main advantage of automated methods is the speed at which they can process the data and for complex data sets this can be extremely useful. However, the deconvolution relies on the algorithms within the software package, of which little or nothing may be known.

Automated scanning of a manually produced list can provide a compromise between the two, identifying peaks already found in previous samples. Once this has been carried out

manual deconvolution of the remaining peaks and addition of identified components to the list used to search the samples will reduce the manual portion of the processing over time as the size of the list increases. This method would rely on two things; that the same peaks appear at the same retention time, within the set retention time window of the processing package, and that the mass spectrum for each peak is reproducible between samples. Drift in retention time can be due to several factors and can be monitored using a standard mix to assess the reproducibility of the system. This standard mix can also be used to create a set of retention indices [170] to correct for any retention time drift in identified peaks. Maintaining reproducible mass spectra can be difficult in a sample as dynamic as breath, as interference from co-eluting compounds can cause the observed spectrum at a specific time to be different. Only manual deconvolution can guarantee that all peaks are inspected and deconvolved.

6.4.4 Breath compound database

The compounds and their associated meta-data can form the basis of a human breath volatile database. Before setting up a database it is important to outline and define a structure for each entry. They must contain all the relevant information in a manner that allows the users to search for it and be presented with the required information. When building a database of VOCs found in human breath this must be taken into account. For example; if someone wishes to compare their list of VOCs for a specific disease to the database they must be able to search by disease. This search should then present them with a list of compounds found in volunteers with the particular disease and allow them to easily cross-reference their own list against the compounds contained in the database.

Allowing for searching based on the mass spectrum is important as not every compound can be identified. This means that compounds that cannot currently be identified can still be included in the database.

6.5 Conclusions

This work has shown that the described methods can be used to analyse breath samples and identify the individual compounds present. The data presented in this preliminary study show that human breath is both variable and dynamic. The current methods used to identify the compounds observed need improvement mass spectral libraries should be used with caution. Classifying compounds based purely on their mass can cause problems when comparing results from different centres. There are two ways to improve this situation

1. Record the compound in terms of mass and retention index.

By using two different physiochemical properties to classify the compounds cross-comparisons of results will be made easier. It will also be easier to identify compounds which are recurring in different sample sets. The use of retention index instead of retention time means that changes to the retention time caused by changes to the column due to damage or aging will not affect the classification of compounds. Identifying a set of compounds that can be used to create a retention index ladder, such as straight chain hydrocarbons, is important. The move to using retention indices instead of retention time should be the aim for breath research as this would allow cross comparison much easier.

2. Record the mass spectra of the analytes from each sample for comparison.

Electron ionisation (EI) mass spectrometry can cause problems. The fragmentation patterns can cause misidentifications of compounds and a reliance on automated deconvolution and automated searching is not sufficient for this type of data. Recording the mass spectrum of each analyte means that these can be compared to the mass spectra of future samples, thus not relying on general database identifications. As a move in this direction the top six mass ions and their intensities were recorded in this work. These can be compared to future samples to more accurately identify if a compound is recurring in different samples.

Chemical ionisation (CI) mass spectrometry is a softer ionisation method that may be beneficial in breath analysis. While CI is selective, based on whether a molecule will react

with the ions introduced by chemical ionisation, it will produce molecular ions, making identification of analytes easier.

The mass analyser used in this work was a quadrupole ion trap. The benefit of using a quadrupole ion trap is that it can perform multiple fragmentations on a selected ion to accurately identify the compounds of interest. The problems with using a quadrupole mass analyser are the effect of space charging in the trap where a high number of ions are present and the poor mass accuracy compared to other instruments. A time-of-flight (TOF) mass analyser could provide greater mass accuracy. Combined with a softer ionisation source this would allow for more accurate identification of the analytes. Recent work in the use of thermal desorption - electrospray ionisation – time of flight mass spectrometry show [89] that this is possible and should be developed to target the compounds expected to appear in breath samples.

The development of a database of compounds seen in breath would allow researchers to easily cross-reference possible biomarkers with other centres samples. This will do two things, it will allow researchers to easier check their compound lists against compounds already described in breath and it will have the effect of building a large knowledge base of samples and what diseases they are present in. By recording what compounds are present in certain diseases this can be used to check any possible biomarkers of a specific disease against the samples associated metadata to see if it could be caused by any other disease that the sample giver was suffering from.

Overall the methods used have allowed for the identification of the compounds found in the breath samples of lung cancer sufferers. The continuation of this study will allow for a preliminary investigation into possible biomarkers of lung cancer. This work can then be extended using the data collected with the differential mobility spectrometer with an aim to develop a point of care device for the detection of lung cancer.

7 Overall discussion, conclusions and future work

7.1 Overall discussion

This research sought to establish the systems and work-flow for prospecting for volatile organic compound markers in the breath of individuals with lung disease. There were four principal objectives:

- To design and commission a Thermal desorption CG-DMS/MS dual detector instrument to increase the amount of data collected from a single sample.

- To develop a method for preparing three dimensional DMS for chemometric analysis while maintaining the full complexity of the data set.

- To set up and implement a collaborative breath sampling clinical campaign within the National Health Service (NHS)

- To evaluate and assess the initial findings from the collaborative breath sampling study.

The implementation of a dual detector instrument was key to this work. Breath samples obtained from people with lung cancer are extremely hard to obtain, and rightly so due to strict ethical governance of who can interact with them in a clinical setting. Only a limited number of samples may be collected from a volunteer and this means that extracting the maximum amount of information possible from every sample is important. The use of two different detectors, an MS and DMS to analyse each sample increased the amount of data gained from each sample. This approach enabled a fieldable system (DMS) to be tested and characterised for use in an in-clinic operation alongside “gold-standard” mass spectrometric detection. The mass spectrometric data were intended to complement and validate the DMS data.

There were some problems with the system:

- The differences in the two chromatographic columns meant that the retention times of the same components were different for each detector. This means that future data processing will need to establish robust approaches to alignment of data for the two channels.

- High levels of water in the breath samples led to an increased column bleed and as such a higher background noise signal.

Differences in retention time may be accounted for by moving to a retention indices method for both detectors. This would also correct for any retention time differences between data sets from the same detector which is a major problem in data processing. Despite the issues highlighted this system still has the major benefit of providing two data sets per sample and at this initial stage of the research this is a major benefit.

With large, complex data-sets such as those produced from breath samples the data processing methods must be capable of analysing all of the data in a reliable way. Multivariate techniques are seeing an increased use in similar research areas such as metabolomics. The main problem with processing DMS data in this way is that multivariate methods such as PCA or PC-DFA cannot be applied to three dimensional data. The current methods or reducing the dimensionality of the data, such as summing across the compensation voltage axis, do not maintain the full information contained within the data. The Euclidian-based transformation method described can maintain all the information contained within the data set while converting it into a suitable form for application of multivariate techniques.

This method was tested extensively to show that it was capable of distinguishing between the smallest changes expected in a real data set. This method was successful in all of the tests applied, showing a significant improvement over the most recently published method. The testing of the method also highlighted the main issue that needs to be addressed, namely alignment of the data sets. Small changes in an analytes position in the retention

time or compensation voltage axis can cause the problems with the data processing and this needs to be addressed if this method is to be utilised to its' fullest. At the moment the follow up research is exploiting the ubiquitous presence of siloxanes in the negative mode to enable alignment on retention indexing.

This method seeks to bring into line the data-processing techniques, which are currently lagging behind, with the analytical techniques. Allowing for processing of the full data set rather than a reduced set makes full use of the data produced and allows researchers to easily utilise the full data set that a DMS can produce for multivariate processing, something that is currently lacking.

For researchers not based in a clinical environment being able to establish a collaboration within a clinical environment is key. A clinical environment provides access to diseased participants in a controlled location, it provides access to experts who can provide meta-data such as tumour staging and information on the participants medical records. It also provides a familiar location that feels safe for the participants to set them at ease. To establish a collaboration within the National Health Service there are several steps that must be completed. Knowing these steps and planning accordingly is vital as the research cannot progress until they have been completed. External researchers require a Research Passport to work within the NHS, all research involving NHS patients as participants requires ethics approval at both the regional and local level and successfully carrying out the sampling campaign requires careful management as it will require interacting with NHS staff.

This work has described what needs to be done implement a successful collaboration within the NHS. To guide the implementation of a collaboration a flow chart and a Gantt chart are provided detailing the steps taken and the time each step is expected to take. These tools are helpful to managing a successful project helping the project manager to track all of the different requirements and keep the research going. While research is ever changing and a project must change to meet the requirements, managing the aspects that must be completed well will increase the productivity of the work.

The preliminary findings from the described pilot study highlight some key points that must be considered as this research progresses.

7.2 Conclusions

- A dual detector instrument was implemented to increase the amount of data produced for each sample. This set-up will also allow for testing of the utility of a DMS as a detection technology for breath analysis

- A new Euclidian based approach to transforming a DMS data surface while maintaining the full structure of the data set has been developed. This method was shown to maintain all of the information contained within a data set and showed a marked improvement over the current method it was compared with

- The limits of the Euclidian based approach were tested and it was shown to be capable of providing data sets that could be distinguished using multivariate techniques based on differences in the data sets that were similar to those expected in breath samples

- A work flow for a collaboration within the NHS was described along with suggested project management tools, namely a Gantt chart and a work flow diagram detailing steps that must be taken to successfully implement the study.

- A description of the documentation required for external researchers to conduct research within the NHS are detailed and included in the workflow and Gantt charts to highlight that these must be completed in a timely fashion so the research is not held up.

- The initial results from the clinical campaign involving patients referred to the two week wait clinic are discussed. The need for a more rigorous method for both assigning and cataloguing the VOCs found in human breath and the associated meta-data.

- The current method of using the highest hit in a database to assign a compound found in breath samples needs addressing as this is not always accurate. Recording mass spectra along with the identification would allow for better cross-referencing of samples to confirm markers.

7.3 Future work

This research has highlighted several key areas that require further work, these are discussed here.

While EI ionisation methods are standard practice for GC-MS this may not be the most appropriate ionisation method. Chemical ionisation methods which are more likely to provide monoisotopic ions would make identification of VOCs easier, particularly for VOCs that are very similar and produce very similar fragmentation patterns.

Direct analysis of the breath samples or real-time breath analysis would be beneficial as this would shorten the analytical time considerably. The first step towards this has been completed by combining a thermal desorption unit with an electrospray ion mobility mass spectrometry instrument [89]. This dramatically reduced the analytical time required per sample. Fine tuning of the ionisation method may be required depending on the likely biomarkers, but this is a step in the right direction.

The alignment of the MS and DMS data needs to be tackled both individually and jointly. Moving to a retention index method would aid alignment for the mass spectrometry data. This method may also aid alignment of the DMS data with the MS data. The alignment of the DMS data was shown to have the largest negative effect on the data analysis and this needs to be addressed. Methods similar to those used in protein analysis could be considered, but any method used must be carefully controlled so as not to add bias to the data set.

Further development of the DMS technology to provide cheap instruments would be beneficial as this would reduce the costs of any point of care device. Development of injection and ionisation techniques more applicable to the samples would also be beneficial.

Appendix I

Table 1 Peak parameters for set A used in Test 1.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	89	0.2	0.9
2	-0.87	0.2	144	0.2	0.9
3	-3	0.2	210	0.2	0.7
4	2	0.2	210	0.2	0.78
5	0	0.2	180	0.2	0.8
6	6	0.2	153	0.2	0.8
7	-1.5	0.2	220	0.2	0.44
8	-2.3	0.2	234	0.2	0.64
9	0.36	0.2	260	0.2	0.58
10	0.99	0.2	92	0.2	0.82

Table 2 Peak parameters for set B used in test 1.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	89	0.2	0.9
2	-0.87	0.2	144	0.2	0.9
3	-3	0.2	210	0.2	0.7
4	2	0.2	210	0.2	0.78
5	0	0.2	180	0.2	0.8
6	-2.4	0.2	170	0.2	0.65
7	0.97	0.2	200	0.2	0.43
8	0.34	0.2	100	0.2	0.84
9	4.12	0.2	270	0.2	0.32
10	0.99	0.2	165	0.2	0.66

Table 3 Peak parameters for set A used in Test 2.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	88.969	0.2	0.45
2	-1.4	0.2	144.578	0.2	0.3
3	-3.16	0.2	210.313	0.2	0.25
4	2.11	0.2	210.313	0.2	0.4
5	0.35	0.2	179.969	0.2	0.31
6	-2.58	0.2	169.86	0.2	0.28
7	0.94	0.2	200.188	0.2	0.17
8	0.35	0.2	100.094	0.2	0.2
9	-4.33	0.2	270.969	0.2	0.23
10	0.94	0.2	164.813	0.2	0.38

Table 4 Peak parameters for set B used in Test 2.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	88.969	0.2	0.45
2	-1.4	0.2	144.578	0.2	0.3
3	-3.16	0.2	210.313	0.2	0.25
4	2.11	0.2	210.313	0.2	0.4
5	0.35	0.2	179.969	0.2	0.31
6	-2.58	0.2	170.875	0.2	0.28
7	0.94	0.2	200.188	0.2	0.17
8	0.35	0.2	100.094	0.2	0.2
9	4.45	0.2	270.969	0.2	0.23
10	1.53	0.2	164.813	0.2	0.38

Table 5 Peak parameters for set A used in Test 3.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	88.969	0.2	0.21
2	-1.4	0.2	144.578	0.2	0.16
3	-3.16	0.2	210.313	0.2	0.09
4	2.11	0.2	210.313	0.2	0.05
5	0.35	0.2	179.969	0.2	0.12
6	-2.58	0.2	169.86	0.2	0.1
7	0.94	0.2	200.188	0.2	0.2
8	0.35	0.2	100.094	0.2	0.13
9	-4.33	0.2	270.969	0.2	0.19
10	0.94	0.2	164.813	0.2	0.20

Table 6 Peak parameter for set B used in Test 3.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	-0.82	0.2	88.969	0.2	0.21
2	-1.4	0.2	144.578	0.2	0.16
3	-3.16	0.2	210.313	0.2	0.09
4	2.11	0.2	210.313	0.2	0.05
5	0.35	0.2	179.969	0.2	0.12
6	-2.58	0.2	170.875	0.2	0.1
7	0.94	0.2	200.188	0.2	0.2
8	0.35	0.2	100.094	0.2	0.13
9	4.45	0.2	270.969	0.2	0.19
10	1.53	0.2	164.813	0.2	0.20

Table 7 Peak parameters for set C used in Test 3.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	-2.58	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.14
8	3.87	0.2	146.61	0.2	0.21
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 8 Peak parameters for set A used in Test 4.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	-2.58	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.21
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 9 Peak parameters for set B used in Test 4.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	-2.58	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.01
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 10 Peak parameters for set A used in Test 5.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	2.11	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.21
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 11 Peak parameters for set B used in Test 5.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	2.7	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.21
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 12 Peak parameters for set A used in Test 6. Peak 8 included as a reference only as this peak drifts in the data files.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	-2.58	0.2	214.344	0.2	0.01
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.21
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 13 Parameters for peak 8 in each data file of set A used in Test 6.

Sample No.	Cv location	Cv spread	Time base location	Time base spread	Intensity
A1	3.87	0.2	146.61	0.2	0.21
A2	3.28	0.2	147.625	0.2	0.21
A3	3.87	0.2	145.594	0.2	0.21
A4	4.45	0.2	147.625	0.2	0.21
A5	3.28	0.2	146.61	0.2	0.21
A6	4.45	0.2	145.594	0.2	0.21
A7	3.87	0.2	144.578	0.2	0.21
A8	3.87	0.2	146.61	0.2	0.21
A9	3.28	0.2	144.578	0.2	0.21
A10	3.28	0.2	148.625	0.2	0.21
A11	3.87	0.2	146.61	0.2	0.21
A12	4.45	0.2	146.61	0.2	0.21
A13	3.28	0.2	145.594	0.2	0.21
A14	4.45	0.2	146.61	0.2	0.21
A15	3.87	0.2	145.594	0.2	0.21
A16	3.87	0.2	144.578	0.2	0.21
A17	3.28	0.2	146.61	0.2	0.21
A18	3.87	0.2	148.625	0.2	0.21
A19	4.45	0.2	144.578	0.2	0.21
A20	3.87	0.2	146.61	0.2	0.21

Table 14 Peak parameters for set B used in Test 6. Peak 3 included as a reference only as this peak drifts in the data files.

Peak no.	Cv location	Cv spread	Time base location	Time base spread	Intensity
1	3.28	0.2	88.985	0.2	0.15
2	0.35	0.2	115.266	0.2	0.2
3	-2.58	0.2	214.344	0.2	0.19
4	1.53	0.2	214.344	0.2	0.15
5	0.94	0.2	159.75	0.2	0.08
6	2.7	0.2	167.844	0.2	0.06
7	-4.92	0.2	101.11	0.2	0.18
8	3.87	0.2	146.61	0.2	0.01
9	-1.99	0.2	177.953	0.2	0.1
10	-0.82	0.2	251.766	0.2	0.18

Table 15 Parameters for peak 3 in each data file of set B used in Test 6

Sample No.	Cv location	Cv spread	Time base location	Time base spread	Intensity
B1	-2.58	0.2	214.344	0.2	0.19
B2	-2.58	0.2	213.344	0.2	0.19
B3	-3.16	0.2	214.344	0.2	0.19
B4	-2.58	0.2	212.328	0.2	0.19
B5	-2.58	0.2	214.344	0.2	0.19
B6	-3.16	0.2	214.344	0.2	0.19
B7	-3.16	0.2	215.36	0.2	0.19
B8	-1.99	0.2	216.375	0.2	0.19
B9	-1.99	0.2	215.36	0.2	0.19
B10	-2.58	0.2	214.344	0.2	0.19
B11	-2.58	0.2	212.328	0.2	0.19
B12	-1.99	0.2	214.344	0.2	0.19
B13	-3.16	0.2	212.328	0.2	0.19
B14	-1.99	0.2	212.328	0.2	0.19
B15	-1.99	0.2	215.36	0.2	0.19
B16	-2.58	0.2	214.344	0.2	0.19
B17	-3.16	0.2	213.344	0.2	0.19
B18	-3.16	0.2	212.328	0.2	0.19
B19	-2.58	0.2	214.344	0.2	0.19
B20	-2.58	0.2	215.36	0.2	0.19

Appendix II

Code for Euclidian transform programme

```
Option Strict Off

Option Explicit On

Friend Class Form1

	Inherits System.Windows.Forms.Form

	' Written by: Sultan Shair

	' Department of Chemistry

	' Loughborough University

	' LE11 3TU



	Public Sub Form1_Load(ByVal eventSender As System.Object, ByVal eventArgs As System.EventArgs) Handles MyBase.Load



		Me.Show()



	End Sub



	Private Sub Dir1_SelectedIndexChanged(ByVal eventSender As System.Object, ByVal eventArgs As System.EventArgs) Handles Dir1.SelectedIndexChanged

		Label1.Visible = False

	End Sub
```

```
Private Sub Drive1_click()

        Label1.Visible = False

End Sub

Private Sub File1_SelectedIndexChanged(ByVal eventSender As System.Object, ByVal
eventArgs As System.EventArgs) Handles File1.SelectedIndexChanged

        Label1.Visible = False

End Sub



Private Sub Dir1_Change(ByVal eventSender As System.Object, ByVal eventArgs As
System.EventArgs) Handles Dir1.Change

        File1.Path = Dir1.Path

End Sub



Private Sub Drive1_SelectedIndexChanged(ByVal eventSender As System.Object,
ByVal eventArgs As System.EventArgs) Handles Drive1.SelectedIndexChanged

        Dir1.Path = Drive1.Drive

End Sub



Public Sub File1_DoubleClick(ByVal eventSender As System.Object, ByVal eventArgs
As System.EventArgs) Handles File1.DoubleClick

        Dim sFilename As Object

        Dim l, i, g, c, j, ma As Object
```

```
Dim Data(3000, 3000) As Double

Dim RowSize As Double

Dim ColSize As Double

Dim RowCount As Double

Dim ColCount As Double

Dim ex(500000) As Object

Dim m(500000) As Object


Dim ApExcel As Microsoft.Office.Interop.Excel.Application

ApExcel = New Microsoft.Office.Interop.Excel.Application


sFilename = Dir1.Path & "\" & File1.FileName


ApExcel.Workbooks.Open(sFilename)



For g = 1 To 3000


        RowCount = ApExcel.Cells._Default(g, 1)

        If RowCount = 0 Then
```

```
                RowCount = g


                RowCount = g - 1
                GoTo Step1
            End If

            System.Windows.Forms.Application.DoEvents()

        Next g

Step1:
        For c = 1 To 3000
            ColCount = ApExcel.Cells._Default(1, c)
            If ColCount = 0 Then


                ColCount = c
                ColCount = c - 1
                GoTo Step2
            End If
            System.Windows.Forms.Application.DoEvents()
```

```
        Next c
```

Step2:

```
        For j = 1 To ColCount
            For i = 1 To RowCount
                Data(i, j) = ApExcel.Cells._Default(i, j)
                System.Windows.Forms.Application.DoEvents()
            Next i
            System.Windows.Forms.Application.DoEvents()
        Next j


        'mat = RowCount * ColCount
        For i = 1 To RowCount
            For j = 1 To ColCount
                ex(l) = ((i ^ 2) + (j ^ 2)) ^ (1 / 2)
                m(l) = Data(i, j)
                l = l + 1 ' counter
                System.Windows.Forms.Application.DoEvents()
            Next j
```

```
                System.Windows.Forms.Application.DoEvents()

            Next i



            For ma = 1 To l

                FileOpen(2, Dir1.Path & "\" & " Corrected " & File1.FileName,
OpenMode.Append)



                PrintLine(2, ma & "," & m(ma) & "," & ex(ma))

                FileClose(2)

                System.Windows.Forms.Application.DoEvents()

            Next ma

            MsgBox("your data is now fixed")

        End Sub

End Class
```

Appendix III

Detection of Volatile Organic Compounds in the exhaled Breath of Patients referred urgently to the chest clinic: a Pilot Study

Investigators:

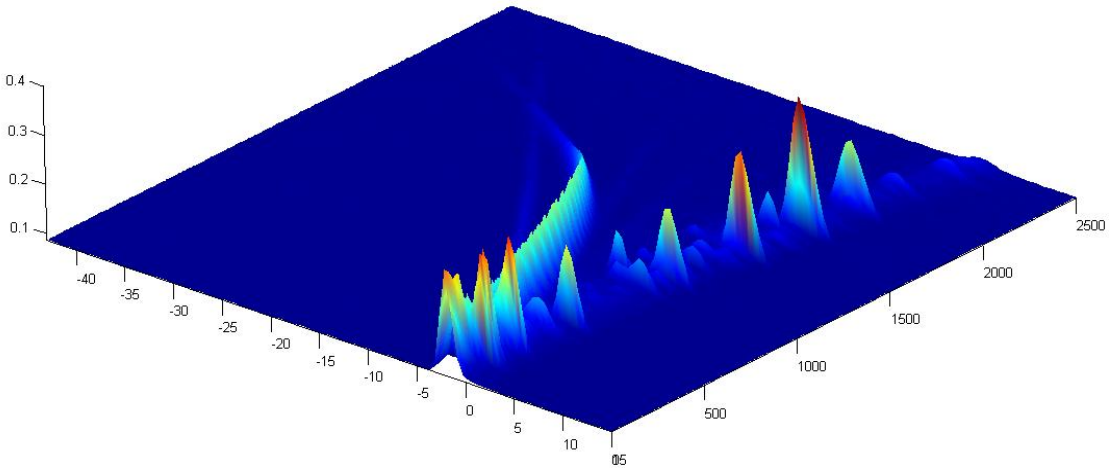
Gavin Blackburn, Ph.D. Research Student, University of Loughborough

Professor Paul Thomas, Professor of Analytical Science University of Loughborough

Dr Stephen Fowler, Lecturer and Honorary Consultant, University of Manchester and Lancashire Teaching Hospitals NHS Trust

Professor Roy Goodacre, Professor of Biological Chemistry, University of Manchester

Professor Ian Wilson, Principal Scientist AstraZeneca, Alderley Park, Cheshire UK.



Background

Respiratory disease can range from simple chest infections to more serious illness. Emphysema, bronchitis, chronic obstructionary pulmonary disease (COPD) and lung cancer can all have very similar initial symptoms. In people with a high risk of serious illness, such as lung cancer, it is important that the disease is detected as early as possible so appropriate treatment can be given. Lung cancer is the most common cause of death from cancer for both men and women. In the UK 24% of all male cancer deaths and 18% of all female cancer deaths are due to this disease[171, 172]. Figure 1, (Reproduced from Reference 1) summarises the distribution of this disease by age and gender within the UK. It is also important to note that lung cancer has one of the lowest survival outcomes of any cancer. Only 25% of patients in England and Wales are alive one year after diagnosis, and only 7% of patients survive for 5 years from the date of diagnosis[173]. For most cases, the only chance of cure comes from surgery, and this is only possible when the cancer is diagnosed before it has spread beyond the primary site. Lung cancer may be diagnosed incidentally when a chest x-ray has been performed for another purpose, for example during a health screen. However in the majority of cases the suspicion is raised by the patient bringing symptoms (for example haemoptysis, chest wall pain, or weight loss) to their GPs' attention. Unfortunately such symptoms often develop after the disease has progressed beyond the point when surgery would be an appropriate treatment.

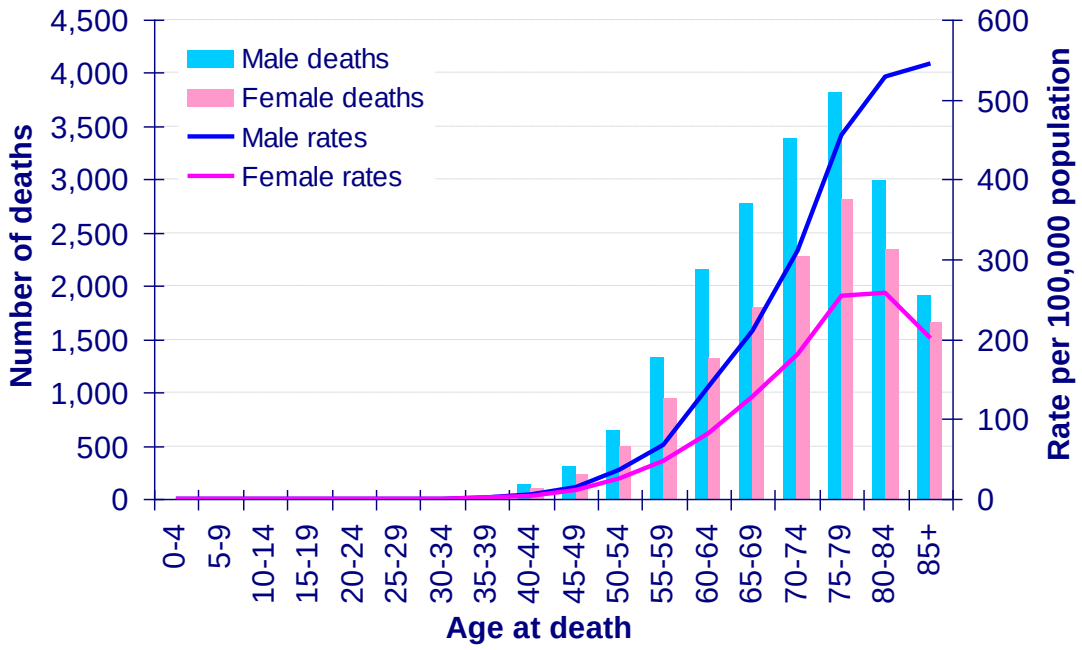


Figure 1 Number of deaths and age-specific mortality rates, lung cancer, by gender, UK, 2005 [172]

In contrast to breast and cervical cancer, there is no agreed screening strategy for lung cancer in the UK. The eventual goal of our research is the delivery of a detection system potentially suitable for use in a screening programme for a range of diseases, including lung cancer. Our hypothesis is that a sensitive non-invasive test that can detect the volatile compounds associated with disease-specific symptoms, such as lung-tumour growth, and metabolism in the exhaled breath of patients will differentiate them from those without disease. The ultimate benefits are clear: sensitive detection of different respiratory diseases before the disease spreads from the primary site enables early

treatment, and increased survival rates for serious illness. Further, more detailed knowledge of how these diseases alter the profile of volatile organic compounds in patients' breath provides new insights into the progression and mechanism of disease. The discovery of volatile markers that are related to therapy and drug regimes will enable objective assessments of the disease state and pave the way for personalised medicine and care programmes to be effectively implemented.

Current diagnostic and screening methods for patients suspected to have lung cancer are expensive, and a typical patient will undergo a minimum of chest radiography (CXR), computer tomography (CT) scanning, and bronchoscopy before a diagnosis is confirmed or excluded.

Modern state-of-the-art chemical detection systems have the theoretical capacity to detect volatile compounds associated with lung cancer, and one such technology , ion mobility spectrometry, and specifically differential mobility spectrometry (DMS) also has the capacity to be miniaturised and used in point-of-care applications.

Previous research in breath analysis using mass spectrometry, reported by Phillips *et. al.*[36], demonstrated the feasibility of differentiating between participants with lung cancer and asymptomatic controls. Further, it was possible to identify the compounds that caused the observed differences. Baumbach and co-workers [30] have reported success in detecting lung cancer from breath samples analysed with ion mobility spectrometry (IMS). Our use of ion mobility technology will improve on this work by using DMS in combination with mass spectrometry (MS) to deliver higher resolution data and enable miniaturised point-of-care systems targeted towards marker compounds for lung cancer to be developed.

Our consortium has conducted an investigation into the detection of Chronic Obstructive Pulmonary Disease (COPD). We have differentiated between healthy non-smoking, healthy smoking and diseased volunteer participants (manuscript in preparation), and we have developed the research protocols to enable us to run a pilot study of patients referred urgently to the chest clinic.

Hypothesis and aims

The research hypothesis is that there is a detectable difference between the volatile organic compounds present in the breath of diseased and healthy participants. The chemical differences can be identified as candidate biomarkers for diagnosed diseases, such as lung cancer, and allow the development of point-of-care systems to specifically monitor these compounds with very high sensitivity.

This study will collect breath samples and chemically analyse them. Using advanced statistical approaches the intent is to establish if it is possible to distinguish between the exhaled breath of participants with diagnosed disease and the exhaled breath of phenotypically matched controls. Further the intention is to determine the chemical identity of any candidate markers that are discovered in this study.

The sampling method is a relatively simple, non-invasive process that has been designed to make sure the participant is comfortable. It has been tested previously with participants suffering from COPD and healthy controls [104], and the results of the early COPD trial are encouraging with some indication that it is possible to differentiate between the breath samples of asymptomatic controls and participants with COPD.

Methods

Setting

A quiet room in Royal Preston Hospital with a medical air supply, a table for breath sampling equipment, three power points, a seat/bed for volunteer participants and a seat for the operator.

Volunteer participants: diseased, healthy.

Volunteers will be recruited from individuals who have been referred by their general practitioner to the chest clinic under the urgent "two-week wait" category that is reserved for patients with possible lung cancer. For example they may have had a suspicious CXR, or symptoms of haemoptysis, or weight loss, and they almost always have a significant smoking history. These patients are investigated using a computerised tomography (CT) scan of their chest and fibre-optic bronchoscopy to further clarify the diagnosis. Where these tests do not *exclude* lung cancer, further tests such as surgical biopsy and positron emission tomography-CT (PET-CT) may be necessary. Following diagnosis a treatment strategy is formulated at the weekly lung cancer multi-disciplinary team meeting and may include any combination of surgery, chemotherapy and radiotherapy, with the aim of cure or palliation of symptoms. Where lung cancer is excluded, patient follow up depends on individual circumstances.

Volunteer participants will be recruited upon referral to the "two-week wait" chest clinic. They will be asked to take part in the study during one of their visits to the chest clinic before diagnosis. Prospective participants will be provided with information about this study by either their consultant, lung cancer nurse specialist or a research nurse, and where they are able and willing informed consent to participate in the study will be obtained. Subjects will have at least 24 hours to decide whether to participate in the study. If possible the breath sample appointment will be made to coincide with their next hospital appointment for consultation or treatment. It is important to note that breath samples must be obtained *before* any treatment or test has been carried out. All participants will receive the same information regarding the study. All participants will be informed that their final diagnosis will be made available to the researchers by sample codes only, no names or addresses will be used.

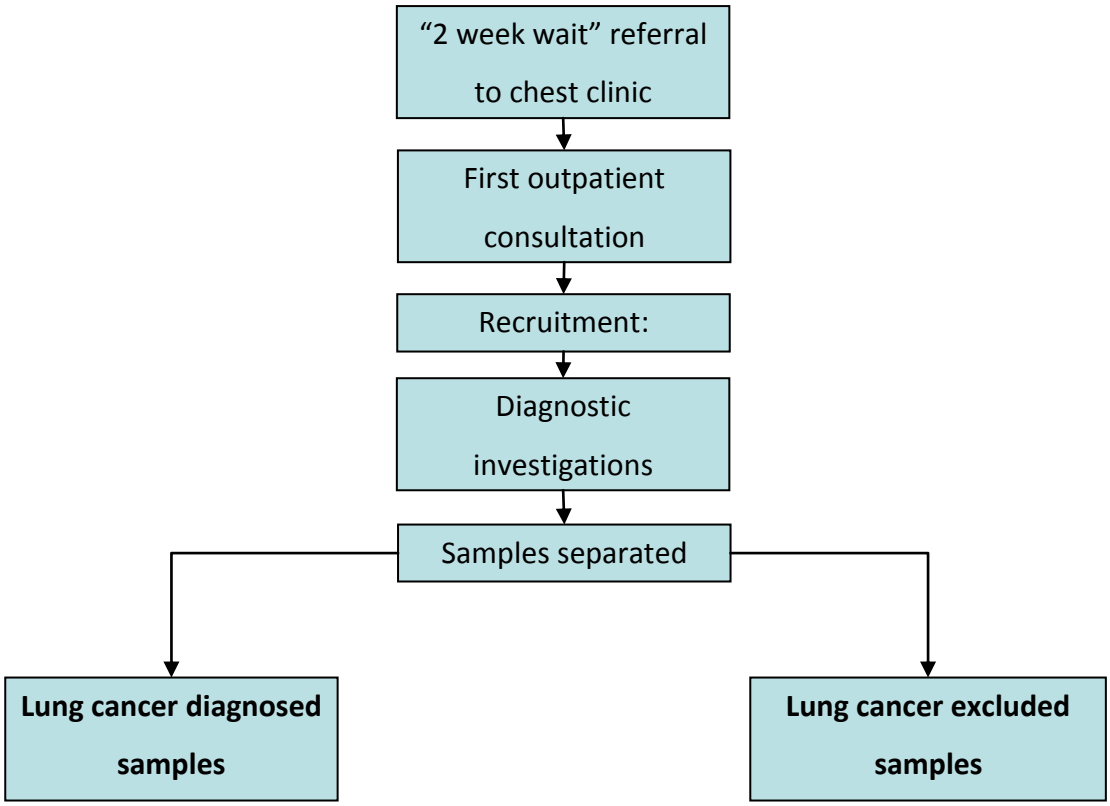


Figure 2 Recruitment pathway for volunteer participants in the proposed study and sample separation

Inclusion criteria

Subjects referred to the chest clinic with possible lung cancer.

Cases: lung disease confirmed

Controls: lung disease excluded

Age > 18yrs

Inclusion criteria may be tailored at the predetermined "halfway" point in the study (after 10 cases and controls have been recruited) to correct for any between-group inequalities in gender or COPD. We know that both these factors may affect the exhaled breath VOC profile from our pilot work in COPD. It is important therefore that the male: female ratio and mean / range of lung function is similar in each group. We plan therefore to review demographics after 10 cases and controls have been recruited, and if there is any gender or lung-function bias, then select participants as appropriate of the basis of gender or lung function for the second half of the study in order to balance numbers.

Exclusion criteria

Co-morbidities that are likely to significantly affect metabolic status. i.e. Diabetes mellitus, chronic renal failure, chronic liver failure

Other active malignancy

Breath collection

An adaptive breath sampler will be used to collect 5 samples of breath from each volunteer. The sampler consists of a full-face mask to which a t-piece, adsorbent tube assembly and a tube attached to a pressure sensor are connected. The t-piece allows for connection of a medical-air line (which is filtered up-line) and a one-way valve. This means that only the filtered medical air will be breathed by the volunteer participant and will effectively eliminate the presence of exogenous compounds due to room air.

The adsorbent tube assembly consists of a capillary held in a reducing ferrule in a Swagelok™ reducing union. This is connected to the mask via a PTFE connection that holds the assembly in place. The absorbent tube is connected to the other side of the union using a bolt and a PTFE ferrule which creates an air-tight connection. The other end of the tube is then connected to the sampling controller that is linked to a computer programme that allows a selected part of the breath profile to be selectively sampled. Figure 3 is a simplified schematic of the sampler, for a more detailed discussion see ref 6.

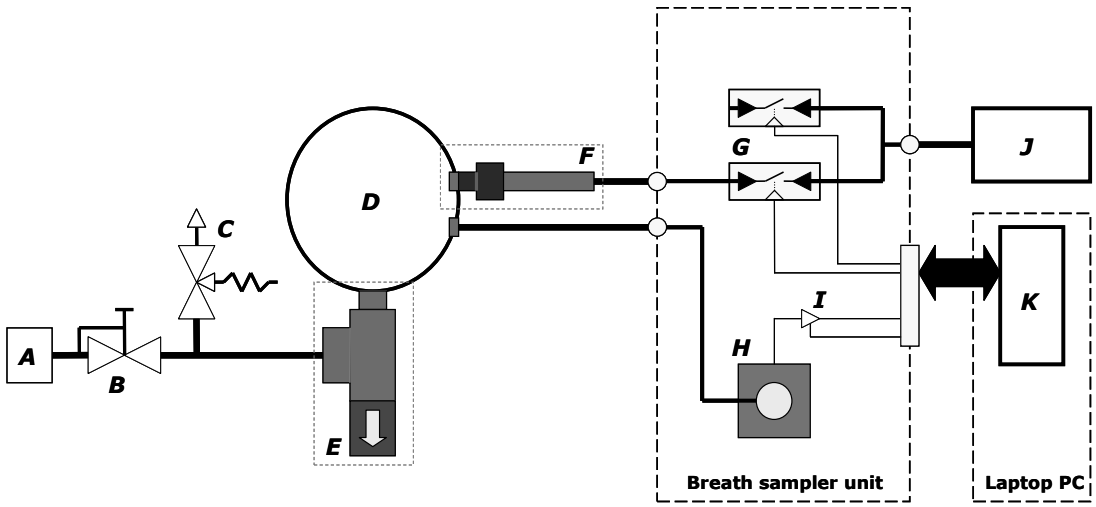


Figure 3 Schematic of breath sampler. A: purified air supply, B: pressure regulator, C: pressure relief valve, D: breathing mask, E: mask air supply assembly, F: adsorbent sampler assembly, G: balanced and matched micro-control valves, H: pressure transducer connected to the mask, I: precision air sampling pump, J: laptop computer

Sampling requires the volunteer participant to sit quietly for about 30 minutes while wearing a full-face mask and breathing filtered air. The sampled portions of breath are collected in tubes containing adsorbents which are attached to the mask and changed periodically (after 2.5 l exhaled air has been sampled). The sampling of the breath is controlled from a laptop running a programme which measures the pressure change within a tube connected to the mask. This allows the system to

detect what portion of the breath that is sampled and allows us to target air from early or late expiration (approximating to the proximal or distal airways) as required. There are several factors that can add bias to the study and these will need to be taken into account. For example, diet, smoking and medication could affect the results. To help us control for this, participants will be asked to refrain from smoking and eating/drinking for 2 hours before the procedure is carried out (unflavoured water will be allowed).

Urine Samples

Urine samples will also be obtained from the participants just before or after a breath sample is taken. These samples will be shipped to AstraZeneca's Alderley Park Research Laboratory where they will be stored, processed and analysed. The complementary data available from urine will be particularly helpful in establishing whether responses seen in breath are due to systemic or local (lung) effects. Note that the AstraZeneca research facility at Alderley Park is covered under the Human Tissue Act and all samples will be stored in a secure freezer on an anonymous basis with access limited to only those people who are directly related with the study.

Sample processing

The breath samples will be analysed using a thermal desorption – gas chromatograph – mass spectrometer/differential mobility spectrometer (TD-GC-MS/DMS). The basic operation of this system is to allow us to detect all the different volatile chemicals in a breath sample. The MS allows us to identify single components and look for differences in peoples' breath. It is an extremely powerful and useful tool and will aid us in identifying potential biomarkers. The data it generates is displayed in two forms; a chromatogram and a mass spectrum. An example chromatogram is displayed below in Figure 4.

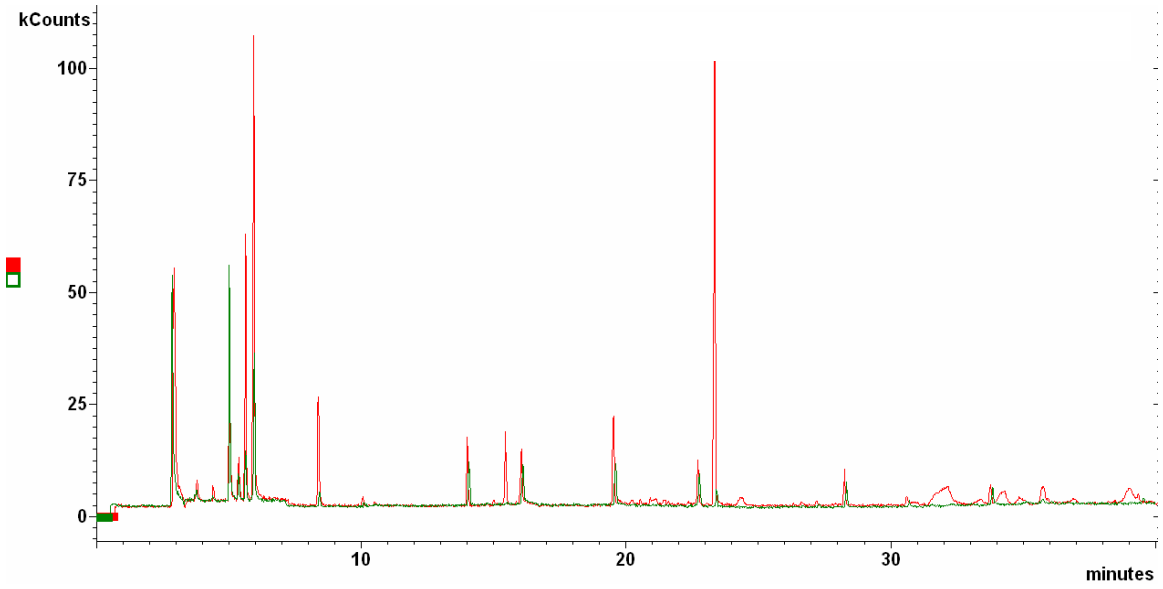


Figure 4 Example of chromatographic data.

The chromatogram above, in Figure 4 allows the relative concentrations of the different volatile components of exhaled breath to be estimated. Each point on the chromatogram is built up from

integrated mass spectra, which may be used to identify the compound. An example is given in Figure 5, a mass spectrum of decane. The x-axis is the mass-to-charge ratio (m/z) and the y-axis shows the relative abundance of the ion fragments yielded by decane to the ionisation source of the mass spectrometer.

The differential mobility spectrometer generates a three-dimensional matrix that can be used like a fingerprint for a disease (see title page for an example of DMS output). The data from this instrument will be used in the statistical analysis to investigate differences between the healthy and diseased groups. DMS has several important benefits giving it the potential to be developed into a desktop detection device for point-of-care applications: it is smaller and easier to maintain than a mass spectrometer, and can also be optimised to detect specific compounds enabling specific markers to be detected with very high sensitivity and selectivity.

Data analysis

All data will be treated as confidential and rendered anonymous at the point of sampling. Data analysis will be carried out under the supervision of Prof Roy Goodacre.

The data flow that we shall use for the analysis of the VOC profiles is detailed in Figure 5, and follows the accepted practices laid out by the *Metabolomics Standards Initiative* [174].

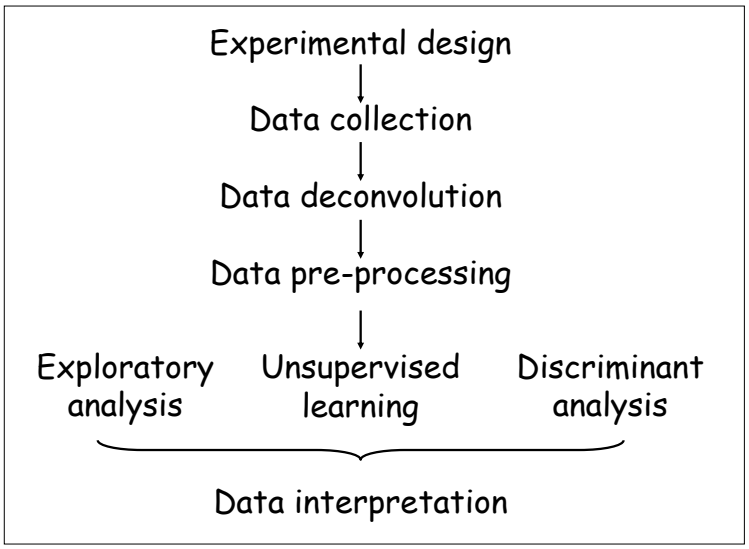


Figure 5 Data processing / analysis flow chart.

Briefly

(1) Data deconvolution and pre-processing will be performed following data collection. Deconvolution will be performed using the instrument manufacturer's software so that VOC peak lists are generated. Pre-processing will depend on the data structure and so will not be explicitly detailed here.

(2) VOC profile reproducibility will be tested using (a) rank testing based on pair-wise comparison of the profiles generated from the same individual and those between different individuals, and (b) principal components-canonical variates analysis (PC-CVA) calibrated with a priori information on the A-C profiles from each individual, 90% chi-sqrd tolerance regions will be constructed and the D profile projected into this PC-CVA space

(3) Unsupervised learning of principal components analysis (PCA) and hierarchical cluster analysis (HCA) will be used to uncover any natural relationship within the samples, this may show groups that are related to disease phenotype or any of the clinical metadata, and is also useful for detecting any outliers (both biological and analytical).

(4) Discriminant analysis using partial least squares with bootstrap methodology will be used to test the hypothesis that VOC profiles can stratify patients according to:

a. VOC profiles of lung cancer versus non-lung cancer controls

b. VOC profiles of small cell versus non-small cell lung cancer

c. VOC profiles of lung cancer versus different lung disease

(5) Individual datapoints from the chromatogram will be compared between groups using ANOVA and ROC analysis, to identify potentially discriminating compounds

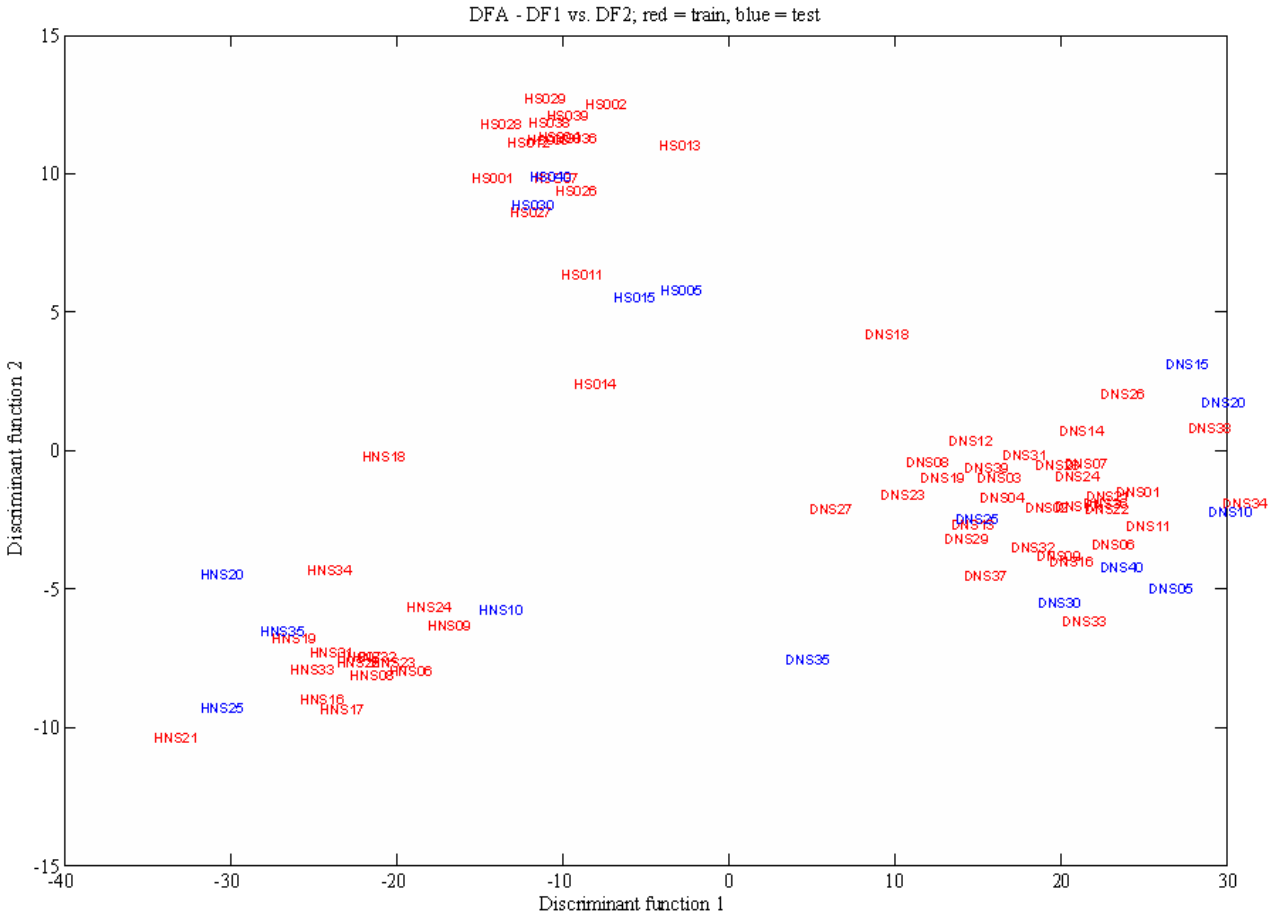


Figure 6 PC-DFA analysis example. HNS = healthy non-smoker, HS = healthy smoker, DNS = diseased non-smoker.

Summary

The breath sampling, analytical approaches and data processing strategies summarised in this application have been awarded a diploma by the International Association of Breath Research for research in breath sampling and analysis, (Prague September 2006). Preliminary data and results from pilot studies with COPD indicate that this is a sensitive approach capable of identifying volatile markers in breath that may be able to discriminate the presence of lung cancer. The proposed study is intended to establish the feasibility of using point-of-care volatile organic compound analysis and detection system to reliably screen breath samples for presumptive testing of lung cancer. If successful this will pave the way for earlier diagnosis with the potential of ultimately increasing survival rates.

Appendix IV

Participant information sheet

Part 1

Invitation to participate

We would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Talk to others about the study if you wish. This part (part 1) tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?

We are testing a new way of detecting changes in the lungs, through measurement of tiny amounts of chemical substances in the breath. These substances arise from all the processes and reactions that occur in the body. We think that analysis of these compounds may be able to give us information about diseases such as lung cancer, and this may lead to a test that can detect the disease early in the future.

Why have I been invited?

You will have been invited because you have been urgently referred to the chest clinic and this makes you part of the group of people we are particularly interested in seeing. Whatever the outcome of your general diagnosis is from the chest clinic, any data collected by us is still useable.

Do I have to take part?

It is up to you to decide. We will describe the study and go through this information sheet, which we will then give to you. We will then ask you to sign a consent form to show you have agreed to take part. You are free to withdraw at any time, giving a reason. This would not affect the standard of care you receive.

What will happen to me if I take part?

If you agree to take part we will arrange a single visit lasting about 45 minutes. If possible we will arrange this for a time when you are coming to the hospital for another reason, for

example for a further clinic visit. At this visit we will first ask you a series of questions relating to factors that can influence the chemical substances in the breath. These will include questions about food, drink and medications recently consumed, smoking, other medical conditions. For the same reasons, you will be asked to refrain from eating, drinking (other than unflavoured water) and smoking for two hours prior to the study visit.

After we have collected this information, we will collect a breath sample. This involves wearing a face mask that is connected to an air supply. You will be asked to breathe normally throughout the test, whilst resting in either a chair or bed. A sampling tube is connected to the mask on which we will collect the sample. 5 samples will be collected in total and the process will take about half an hour.

You will also be asked to provide a urine sample during the study visit. We will analyse this in a similar way to the breath samples, as it may be possible that similar chemical changes can be detected in the urine as well.

None of your usual medication or any proposed treatment will be changed or delayed if you participate in this trial.

What will I have to do?

As mentioned above we need you to not smoke, eat or drink (although you are permitted to drink unflavoured water) for two hours prior to the study visit. In addition we ask you not to use perfumed products (which include not only perfume, but also deodorants, after-shaves), or any facial products (such as make-up or creams (scented or un-scented)) on the day of the study visit.

What are the possible disadvantages and risks of taking part?

We do not anticipate any risk from taking part in this study. Some participants may experience slight initial discomfort wearing the mask, although we have tested this previously in patients with chronic bronchitis and emphysema without any difficulties. Some patients experienced slight claustrophobia when first wearing the mask but this was in very few cases and was not enough to cause anxiety.

What are the possible benefits of taking part?

There will be no direct benefits to you from taking part in this research, The aim is to hopefully develop a test that may be used in future to detect lung cancer and other respiratory diseases in patients with these conditions.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. You may withdraw from the study at any time; either before or during. The detailed information on this is given in Part 2.

Will my taking part in the study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

What will happen if I don't want to carry on with the study?

If you decide to withdraw from the study before or during the study visit this will not affect your clinical treatment at all. You may elect to withdraw from the whole study, or to contribute only to a part of the study (for example if you wish to give a breath sample, but not a urine sample, or *vice versa*), in which case we will use only the sample(s) you have agreed to provide.

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (contact number 01772 523237). You may ask a researcher to clarify anything about the study at any time, both before and during sampling. If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital.

In the event that something does go wrong and you are harmed during the research due to someone's negligence then you may have grounds for a legal action for compensation against Lancashire Teaching Hospitals NHS Trust but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you (if appropriate).

Will my taking part in this study be kept confidential?

All information which is collected about you during the course of the research will be kept strictly confidential, and any information about you which leaves the hospital will have your name and address removed so that you cannot be recognised.

Involvement of the General Practitioner

With your consent, your General Practitioner will be informed of your participation in this study.

What will happen to the samples I give?

The breath samples are collected in small tubes that will be sent to the Department of Chemistry at Loughborough University for processing. Your name will be removed, and the samples identified only with a code number and initials. The urine sample will be processed here at Preston Hospital and sent for analysis to AstraZeneca laboratories at Alderley Edge. Your name will also be removed from these samples which will be identified by code number and initials only. No genetic testing will be performed on any of your samples.

What will happen to the results of the research study?

It is intended to publish the results of this study in a scientific journal. Data generated will be kept at both Astra Zeneca laboratories at Alderly Edge and the Department of Chemistry at Loughborough University. Your name and address will not be kept at either of these locations and the data will only be made available to people who are directly involved with the study.

Who is organising and funding the research?

This study is a collaboration between Lancashire Teaching Hospitals, the Universities of Manchester and Loughborough, and AstraZeneca, and has been planned and funded jointly by these institutions.

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the regional Research Ethic Committee to protect your safety, rights, wellbeing and dignity.

Who should I contact for further information?

Dr Stephen Fowler
Lecturer and Honorary Consultant in Respiratory Medicine
Lancashire Teaching Hospitals NHS Trust
Royal Preston Hospital
Fulwood
Preston
PR2 8BP
Tel: 01772 523237

Thank you for taking time to read this information sheet and considering helping us with our research. Please retain a copy for your reference.

Appendix V

Volunteer Questionnaire

Detection of volatile organic compounds in the exhaled breath of patients referred to the chest clinic at (state location): a pilot study

We will be measuring the chemicals in your breath. Some of these will have come from your diet, lifestyle it helps us a lot if you are able to go through the questions below and provide us with as much information as you can about: make-up, diet and activity All answers will be treated in confidence, with none of your answers being released outside of the research team and no-one other than the principal investigators will be able to link your name to your answers.

Volunteer Questionnaire:

Date:

Patient details	
Name	
Patient Reference	
Age	
Sex	
Ex/current Smoker?	
Pack Years	
Average Calories per day	

Today	
Are you wearing any make-up? Have you used any facial products?	
Are you wearing any perfume?	
Have you used any toothpastes/ mouthwash?	
Have you eaten at all prior to procedure? If yes, what have you eaten?	
Have you drunk anything other than plain water? If yes, what?	
Last bowel movement?	**0-2 hours 2-6 hours 6+hours**

Exposure within the last 72 hours?	**Place of exposure**
Petrol	
Dry cleaning	
Alcohol	

Cigarette Smoke	
Freshly painted surfaces	

This week			
Average alcohol intake (units):			
Portions of:	**Fish**	**Cheese**	**Fruit**
Caffeinated drinks(cups/cans):	**Tea**	**Coffee**	**Soft drinks**

<u>Researcher use only:</u>

Sampling equipment:				
Tube 1:	**Tube 2:**	**Tube 3:**	**Tube 4:**	**Tube 5:**
Mask No.:				
Capillary Number:				

Notes:

Appendix VI

Designing and implementing a collaborative breath sampling clinical campaign within the National Health Service (NHS)

Introduction

Breath sampling research can take place in many different locations depending on its aim. Where the subjects are healthy volunteers they can normally be recruited through the research institute where the sample collection and analysis can be performed in the same place. While healthy volunteers and others with less debilitating symptoms can reasonably be asked to attend a laboratory for sampling, people with serious conditions and perhaps with much reduced life expectancy cannot be asked.

Ethical approval for studies that seek to ask participants with extremely serious clinical conditions to travel to a research facility will not, nor should it, be granted ethics This means that for volunteers in this group the recruitment and sampling must take place at a location convenient to them, and in such a manner that does not put under any undue stress on them. The best approach is to locate the research facility into the hospital or clinical infrastructure. Researchers in these locations have access to clinics for volunteer recruitment and the necessary analytical instrumentation to process the samples rapidly.

However breath profiling is not a mainstream clinical approach and at the moment researchers do not have access to both volunteers and the analytical instrumentation. Multiple site collaboration allows such studies to proceed. Within the UK research, collaboration between an NHS facility and a research team enables techniques of potential benefit to clinicians in the future to be developed without the expense and risk of disruption the clinical operation to a significant extent. There are, however, many steps and processes that must be accomplished to establish the research concepts, infrastructure and access to the clinics to meet with volunteer participants to take samples. All of this activity needs to be planned and managed from the outset of the proposed research collaboration.

This chapter describes the activity and preparation needed to ensure a clinical study can take place. An important element in this is the documentation that must be generated to accompany such activities. The project management aspects are also discussed along with a timeline, in the form of a Gantt chart.

Initial discussion with clinical lead

The clinical lead should be a senior and experience member of the clinical staff where the proposed research is envisaged to occur. Initial discussions with the clinical lead address the proposed research and enable the clinical expert to assess the proposed sampling campaign. After which the clinical lead will also lead the external researchers through the:

- Recruitment of volunteer participants;
- Location for sampling within the NHS site;
- Possibility of NHS staff assisting during the campaign and;
- The amount of access the external researchers will be granted to the volunteers.

During these initial discussions the researchers should develop a plan that describes and justifies:

- How many samples are needed from each volunteer participant;
- How many volunteer participants are needed;
- How the samples are to be analysed:
- How the data are to be processed and;
- What, if any, metadata from the participant's medical records will be required.

Study plan

It is essential to develop a detailed plan of the whole study, from initial recruitment through sampling, analysis and data processing. This study plan will form the basis of any research protocol that will be prepared. Most aspects of the study plan should have been discussed at the initial meetings with the clinical lead. There are several important questions that this plan should seek to address, these are:

- How will volunteers be recruited and where will they be recruited from?
- Who will take the volunteers consent?
- Where will samples be taken?
- How many samples will be taken and how long will a volunteers visit last?
- What are the logistics for transporting/delivering the samples to the laboratory?
- How will the samples be analysed?
- How will the data be processed?
- Who will have access to sensitive volunteer information?

By answering all these questions the major points for the research protocol will be covered. There may not be answers to some of these questions, such as where sampling will occur, but the answers should seek to provide specifications so identification of suitable sites and procedures at a later stage is straightforward. In this work for example the sampling site was specified as:

> "A quiet room with space for a researcher and a volunteer, with a medical air line to supply the breath sampler and three 13 A 3-pin sockets for the sampling equipment. A comfortable chair/bed for the volunteer to use during sampling and a table to set up the sampling equipment is also needed."

Once such a specification has been established in the study plan identifying possible sites becomes more straightforward. Later on in the study it became clear that most volunteer

participants had at least one person accompanying them, and so their comfort and welfare also needed to be taken account of.

Obviously the study plans evolve and develop and specification may change. For example in this project the recruitment procedure was modified to allow the participant to waive the 24 hour period of reflection if the participant was happy to do this; collecting samples on their initial clinical visit was found to be more convenient for them than planning a follow-up visit. This was because the volunteers attended the chest clinic in the morning, underwent diagnostic tests from mid morning to early afternoon and then saw their consultant for a follow up later on in the afternoon. This meant that the volunteers had a large period of time where they were at the clinic and could give samples. While this was mentioned in the ethics application forms, it was submitted as a change as this became the norm and it was felt by the researchers that the ethics board should be informed of this. Changes like this are to be expected and the plan should evolve with the study, but having an initial plan can help highlight these problems early on.

In summary all activities should be fully planned and understood along with the expected research outcomes. Once a study plan has been discussed and finalised it is time to write the research protocol.

Ethics

When a study requires interaction with volunteers from a clinical setting, an ethics application must be made to the relevant NHS regional ethics board. Making an ethics application means submitting a proposal to a NHS regional ethics board, who will then review and make a decision as to whether the research is worthwhile. The members of the ethics board will take into account potential benefits and risks, and whether all the expected ethical standards have been met.

There are two documents that are required for an ethics application, the research protocol and the online ethics application form. Supplementary material such as participant

questionnaires should be included in the research protocol. This section describes these and reviews the online ethics forms that must be completed so the application can be made.

Research protocol

The protocol must make the case for the research; discuss the relevance of, and the need for, the study. This is the document on which the research will be judged, and a decision will be made as to whether the research meets the required standards. There are many factors that will be considered when deciding if the research reaches the standard required for ethical approval including the following five general areas;

It will take into account the volunteers who will be asked to participate and they must not be put under any unnecessary stress,

Any potential benefits from the study must be outlined,

What access, if any, to volunteers medical records must be outlined, along with who will have access to the information and what will be done to keep volunteer involvement confidential,

What the data will be used for, what processing it will undergo and what output is expected,

What, if any, analysis of samples containing volunteer DNA will be undertaken, where any such samples will be stored and what the data will be used for.

The ethics committee will often provide the researchers with constructive feedback to assist them in undertaking better designed and more skilfully executed work to higher ethical standard. For example:

> During the writing of the protocol for the study in two week wait patients the aim of the study was altered from investigating volunteers with lung cancer to investigating volunteers urgently referred to the two week wait clinic. This change was because the point of sampling could occur before the volunteer had received a diagnosis from a

clinician and any mention of lung cancer in the study documents would cause undue stress. This is an example of where a study needs to be changed so it can be ethically conducted. Volunteers with lung cancer would still be included in the study, as volunteers could be selectively recruited based on the clinicians' prior knowledge, but no undue or unnecessary stress would be caused in any volunteers. Submitted with the research protocol should be any patient information forms (Appendix IV) and questionnaires (Appendix V) and any other forms that are required. The Research Protocol for this work is included in Appendix III.

Online ethics forms

Once a research protocol has been finished, the online ethics application forms must be completed and submitted, along with the research protocol, for ethical approval. These can be found at www.myresearchproject.org.uk. Much of the information required for these forms can be extracted from the research protocol. The bulk of the forms can be completed by the external researcher, with the clinical lead completing any questions relating to the NHS location and clinical aspects of the trial.

Ethics approval meeting

Once the Ethics documentation has been submitted, a meeting will be scheduled by the regional ethics board where the research will be discussed between the members of the board, and any recommendations can be made. The researchers are invited to the meeting so they can answer any questions the panel may have regarding the study. Once this meeting has taken place the board will make their decision and pass their recommendations on the researchers. The board may grant the study approval after the first meeting or the recommendations may require changes to be made to the protocol. If changes are required the ethics application may need to be re-submitted and reconsidered. Sometimes the changes are minor and the board requires written notification that they have been made before they grant approval.

Once approval has been granted by the regional board, the documentation can be passed on to the location ethics board.

Local ethics approval meeting

Once the regional NHS trust ethics board has granted the study ethical approval, the documentation must be past to the NHS locations local ethics board for approval. The local ethics board will review the applications and make recommendations for any changes they feel are necessary to the protocol. With these changes made the ethical approval process is completed.

Researcher documentation

There are several pieces of documentation required for different stages of the study. The bulk of this documentation is required for any researcher who wishes to obtain an honorary contract with an NHS location. An honorary contract is required for researchers who wish to participate in taking samples from volunteers.

Curricula Vitae

The chief investigator, research student supervisor and research student must all submit curriculum vitae along with the ethics application documents.

NHS research passport

Outside researchers are required to hold a research passport if they wish to work within an NHS location and they fall within the pre-defined groups set out in the literature. Documentation is available from the National Institute for Health Research on who requires a research passport and how to apply for one [158]. The research passport allows researchers to obtain an honorary contract to allow them to carry out research within NHS locations. It was implemented so that external researchers could move from one NHS site to another and obtain an honorary contract without having to fill out the same paperwork at each location.

Research passport documents

There is a standard form to be filled out and a set of checks that are carried out on every researcher who applies for a research passport. This means that a researcher can move between locations and start new studies without the need for the same checks being carried out each time. As well as the forms that must be submitted there are two independent checks that may be required depending on the work being carried out. These are an occupational health check and a criminal records bureau check.

Occupational health

An occupational health check is required in some circumstances, depending on what research is being carried out. This check is undergone at the researchers' place of employment, not within the NHS. The occupational health check will make sure the researcher is capable of carrying out the work within the hospital. It will also make sure that the researchers' immunizations and vaccinations are up to date so they are not a risk to the volunteers they must interact with. This protects both the researcher and the NHS location where they will be working as it provides evidence that they are fit to work. The occupational health department will require the research protocol so they can assess what work will be undertaken within the NHS and what checks must be made. If any amendments are made to the protocol that may require additional checks, the researcher should inform the occupational health department. Once the checks have been carried out and the researcher has been cleared to work, the forms must be signed and passed back to the department coordinating the application for the research passport.

Criminal records bureau (CRB) check

The criminal records bureau is "…an executive agency of the Home Office set-up to help organisations make safer recruitment decisions [159]". A full and detailed description of the checks that can be carried out, a list of the data sources they use and their code of practice can be found on their website [160]. The CRB check is carried out to make sure the researcher is suitable to carry out certain aspects of the study. This may include interacting

with vulnerable patients such as children or vulnerable adults, having access to patient medical records or operating within patient areas of an NHS hospital. The CRB check must be sponsored by the researchers' employer. The forms that are required will be issued to the researcher, who must fill them out and return them to be processed and sent to the CRB. Notification of the result of the check will be sent to the researchers' address, this may take 2-3 months and this time should be allowed when planning the clinical study. Once the check results have been received a copy of them should be sent to the department who are coordinating the application for the research passport.

Honorary contract

Once the research passport has been completed and signed off, the researcher can apply for an honorary contract with the NHS location where the research will be carried out. While the research passport is required for the researcher to apply for an honorary contract, the honorary contract is required if the researcher wishes to work within the NHS. Once an honorary contract has been granted, the researcher is allowed to interact with volunteers without direct supervision from the clinical staff. This means that the researcher can operate within the hospital without interrupting the clinical staff when they are working.

Clinical location set up

Identification of sampling site

The identification of a suitable sampling site is discussed during the initial research meetings with the clinical lead. The description of a suitable location and the required elements will have been set out in the study plan to specify an appropriate site. It is helpful to note that there are several things that need to be taken into account when locating a site within a large organisation such as an NHS hospital. As many of the possible locations could be in different departments, discussion with the relevant staff about availability of sites will help eliminate anything unsuitable. Depending on the demographic of the volunteers, it may be necessary to limit the site to a specific area of the hospital. Also, the required services, particularly the medical air line, are likely to be in very specific, and possibly limited, places.

The clinical lead should be able to identify any suitable possibilities and once this has been done an inspection of each site and discussion between the clinical lead and the external researchers should limit the sites down to one or two. These sites can then be discussed to **confirm availability and a specific site can be chosen.**

Installation and testing of sampling equipment at site

Once a site has been chosen it is then necessary to install and test the sampling equipment. In this study all the sampling equipment could be transported to and from the sampling location, meaning that secure storage of the equipment was not an organisational factor problem. Tests were conducted on-site to ensure that all the equipment functioned properly and the site was comfortable for the volunteers and the researchers. On site testing allowed any mechanical faults with the sampling equipment caused by transport from the laboratory to be identified and fixed. A major element of the breath sampling equipment is the medical air supply pressure required to pressurise the external air filters. Medical air is supplied via a generator within NHS hospitals and it must be regulated to a maximum pressure of four bars when being used to supply air to a face mask. This is an important and small detail that reflects the need for careful and detailed assessment of the site.

Volunteer recruitment

The process of volunteer recruitment will have been set out in the ethics protocol. This will have been reviewed by the ethics board and approval given. This process should be discussed with the relevant clinical staff. In the case of this research, the clinicians who ran the two week wait clinic and a research nurse recruited the volunteers. The role of research nurses in this aspect is very important as they generally are able to achieve effective rapport with the population of potential volunteers and answer questions in a straight forward and empathetic way. Clearly, it is important to provide the clinical staff with straightforward briefings about the research and effective summaries of the participant information sheets. Researchers should ensure that they are available to discuss the ethics and associated protocol, the volunteer information sheet (Appendix IV) and to answer any questions with

the clinical team. The researchers need to ensure that everyone involved in the study knows what the aims are and what is required for it to be successful. Staff changes are normal in many clinical settings so the researchers should plan to do this regularly.

Infection control

Different clinical settings have different local rules regarding infection control. It is essential that the research team inform themselves of the local rules and abide them. The preparation of the sampling equipment necessarily involves cleaning and disinfection of the components that are reused. Appropriate training, supervision and audit are required to ensure that this essential element is correctly followed.

Breath sampling

Before the study can start, the breath sampling procedure needs to be thoroughly tested. This will highlight logistical requirements as well as test the sampling equipment again. The whole process should be tested from sample transport, sample acquisition to sample recovery and analysis. (This means that if it is planned to post the samples from the clinical location to the laboratory this should be done, even if the researcher is currently travelling between locations during the set-up.) These tests also allow clinical staff to observe the process, provide suggestions and observations and ask further questions that occur to them.

It may be that clinical staff will take over the sampling and so the initial sampling run through provides a good starting point for training. Once the run through has been completed and the process and logistics have been tested and are satisfactory, the study can be begun.

Sample analysis

The analytical protocol must be validated before the study starts, normally before the study is discussed with the lead clinician, although further development of the analysis method can continue until the samples have started to be collected. The samples must be

catalogued so that they can be traced back to the volunteer who gave them, but should not include any personal information.

Sample storage and transport should be considered. For this work, as no storage study of adsorbent trap breath samples has been carried out, the samples were transported in "freezer packs" and stored at 4 degrees Celsius. All samples were analysed within 72 hours of sampling.

All the samples should be analysed using the same method of analysis so that they are directly comparable. This means that the instrumentation must be established in the laboratory and maintained to a high standard during the period of the study. In this work all of the samples were analysed on a thermal desorption gas chromatography mass spectrometry/differential mobility spectrometry platform. This provided two different sets of data for each sample, so gaining the maximum amount of information for each sample collected.

Flow diagram and Gantt chart

Using a flow diagram illustrates the relationships between the steps that are needed to set up the clinical study. It gives researchers an easy reference to check where they stand in each individual task, and what needs to be done to complete them. As a quick reference tool it can keep the work focused on the required tasks and make sure everyone in the team knows what is required. As a tool it gives a project manager a quick guide as to what is required of each team member at each stage of the project and means that the next steps can easily be moved on to.

A useful tool for planning a study is a Gantt chart. A Gantt chart allows a team to see what steps need to be taken and when along a project timeline. It allows them to see what should be being worked on at any point and what goals need to be achieved at certain times. It shows what tasks need to be completed before other tasks can be started and assigns the workload for each person working on the project. Like the flow diagram it shows any

dependencies for each task and can also show the percentage complete of each task during the process.

Figure 5-1 and Figure 5-2 show a flow diagram and an “idealised” Gantt chart for setting up a collaborative breath sampling study within the NHS.

Appendix VII

Table 1 Compounds found in the breath of participant CV-08

	Compound	M/Z						observed in
1.438	Benzene	77.9	76.9	49.9	50.9	51.9	73.9	BC1, BC2
Intensity		155152	29752	28566	27522	22896	11755	BC1, BC2
1.608	No Hits	72.8	73.8	54.7	56.7	55.6	74.7	BC2
Intensity		39300	32743	10867	10048	6293	5767	BC2
1.683	No Hits	76.9	44.9	72.8	77.8	69.3	56.5	BC2
Intensity		24844	8086	7464	3573	3045	2870	BC2
1.739	No Hits	93.9	92.3	44.9	96.4	91.2	94.9	BC1
Intensity		39459	10395	9632	8603	6920	5143	BC1
1.926	No Hits	82.4	54.6	80.8	66.9	93.7	79.7	BC1
Intensity		18576	17821	8639	5366	4854	4566	BC1
2.395	Toluene	90.9	91.8	62.9	64.8	92.9	88.9	BC1
Intensity		503719	291569	59234	33341	26039	24945	BC1
2.75	No Hits	93.9	175	176.9	176	93	91.8	BC1
Intensity		28518	17163	13894	8977	8348	8308	BC1
3.003	No Hits	94.4	96.7	44.9	95.6	92.6	76.8	BC1
Intensity		45503	11153	9445	6064	4480	4016	BC1
3.434	No Hits	94.4	72.8	73.7	44.9	54.8	96.7	BC1
Intensity		19075	14738	11432	7432	5287	5201	BC1
4.046	p-xylene	91	105.8	94.9	91.9	77.9	105.1	BC1, BC2
Intensity		113272	24225	10539	9314	8927	8754	BC1, BC2
4.253	o-xylene	91	105.9	105.1	94.9	79	92	BC1
Intensity		104722	50672	16470	13687	10854	8840	BC1
4.422	No Hits	102	75.9	73.9	49.9	63	74.9	BC1
Intensity		19659	4622	3885	3807	3459	3310	BC1
4.792	styrene	104	78	91	103	105.9	50.9	BC1, BC2
Intensity		91752	54005	50443	46519	28467	18711	BC1, BC2
4.851	No Hits	72.9	55.7	54.8	69.7	68.8	96.5	BC2
Intensity		4283	4259	3566	3531	3466	2158	BC2
5.076	No Hits	56.7	55.8	84.5	54.8	73	69.8	BC1
Intensity		11594	4363	4236	3313	2689	2598	BC1
5.962	No Hits	92.9	91	91.9	78.9	151	77	BC1
Intensity		26880	15800	13047	8785	6877	6144	BC1
6.608	No Hits	90.9	149	72.9	91.9	151	119.6	BC1
Intensity		16198	3610	3373	3264	3213	2999	BC1
6.9	2-chloro-acetophenone	104.9	76.9	50.9	94.9	49.9	105.7	BC1, BC2
Intensity		218848	83585	58736	45837	42413	26801	BC1, BC2
7.143	No Hits	94	105	120	66	95	63	BC1
Intensity		11806	10752	5766	5317	4404	4141	BC1
7.378	Phenol	94	66	92.9	64.9	91	63	BC1
Intensity		24405	11671	10418	5520	5493	5034	BC1
7.416	No Hits	94	66	105	92.9	64.9	91	BC1

Intensity		23688	15470	13462	7934	5520	5141	BC1
7.597	Phenol	93.9	66	64.9	62.9	61.9	94.9	BC2
Intensity		99002	43159	17651	9186	6317	6264	BC2
7.668	Benzonitrile	103	75.9	66	93.9	50	75	BC1
Intensity		99088	28782	21178	16121	15349	14613	BC1
9.249	No Hits	56.7	55.7	69.8	54.8	82.7	57.7	BC1, BC2
Intensity		222443	29467	24761	15865	12230	8761	BC1, BC2
9.332	Limonene	92.9	79	67	93.8	67.8	90.9	BC1, BC2
Intensity		34674	27240	27177	22600	16346	15377	BC1, BC2
9.509	Hydrocarbon	56.7	69.8	70.6	55.7	54.8	82.7	BC1, BC2
Intensity		80687	42328	33793	27793	25507	20663	BC1, BC2
9.646	Benzyl Alcohol	78.9	76.9	106.9	107.9	72.8	50.9	BC2
Intensity		16654	5389	4916	4644	3751	3738	BC2
10.075	No Hits	56.8	69.8	70.6	55.8	54.8	82.7	BC1
Intensity		51298	28256	18031	17706	10741	7989	BC1
10.746	2-chloro-acetophenone	104.9	76.9	50.9	94.9	49.9	105.9	BC1, BC2
Intensity		81905	32830	24083	16134	8891	5349	BC1, BC2
12.52	No Hits	81.1	66.9	69.7	95	55.7	56.8	BC1, BC2
Intensity		21390	17263	16023	15415	15236	13817	BC1, BC2
16.813	methylated alkane	80.9	67	67.9	54.7	82.8	81.8	BC1, BC2
Intensity		14768	12693	8715	8025	7997	7841	BC1, BC2
16.79	No Hits	80.9	67	67.9	82.8	54.8	81.9	BC1
Intensity		28517	23597	17643	17127	15827	14212	BC1
20.296	No Hits	117.1	90	141.1	142.1	88.9	62.9	BC1
Intensity		33308	15631	14894	14641	12439	7540	BC1
23.958	No Hits	91	119	105	92.9	72.9	93.9	BC1
Intensity		10628	9673	7094	6901	6703	6370	BC1
28.027	No Hits	107	93	69	79	151	91	BC1
Intensity		17248	11602	10275	7018	6169	5504	BC1

Table 2 Compounds found in the breath of participant CV-14

	Compound			**M/Z**				**Observed in**
1.403	Benzene	78.1	50	51	77	55.8	52	JL1, JL2
Intensity		487927	118291	92466	82742	62504	62173	JL1, JL2
1.548	No Hits	40.9	43	73	74	54.9	69.7	JL2
Intensity		59973	48989	44245	38604	35878	35253	JL2
1.611	No Hits	40.8	42.9	77.1	39.9	69.7	70.5	JL2
Intensity		52226	46437	46075	20341	19718	18744	JL2
1.728	No Hits	44.9	43	88.7	40.8	42	60.9	JL2
Intensity		115876	94660	40881	23842	17396	15064	JL2
1.837	No Hits	82.7	54.9	40.9	42.9	55.8	41.8	JL2
Intensity		93209	65141	46832	29850	27771	22963	JL2

1.927	No Hits	45	39.9	40.9	87.9	43	73	JL2
Intensity		23216	18830	17718	16017	15876	10046	JL2
1.982	No Hits	79.1	40.9	42.9	95	39.9	77.1	JL2
Intensity		29255	18115	17467	15365	14174	13774	JL2
2.272	Toluene	91.1	92.1	63.1	65	50.1	51.1	JL1, JL2
Intensity		1009669	700375	168262	83594	75943	75217	JL1, JL2
2.434	No Hits	96.6	40.9	54.8	40	68.9	91.2	JL1, JL2
Intensity		87986	35749	34370	20063	18598	17817	JL1, JL2
2.706	No Hits	40.8	42.6	96.7	84.2	55	41.8	JL1, JL2
Intensity		67152	45187	41058	30729	26549	25294	JL1, JL2
2.742	Hexanal	40.8	55.8	42.9	43.9	71.6	54.8	JL1, JL2
Intensity		84767	75661	48478	47649	32883	26555	JL1, JL2
2.824	1,1-dimethyl-cyclohexane	96.6	40.8	54.9	39.9	43	55.7	JL2
Intensity		65551	28365	26163	14654	12145	10897	JL2
2.841	No Hits	96.1	42.9	95.1	54.9	40.9	91.1	JL1
Intensity		34613	29048	26442	24533	17384	9672	JL1
2.877	No Hits	94.3	95	43	93.7	96.3	44.9	JL1
Intensity		49958	33468	30832	29989	17903	16789	JL1
2.958		43	42	94.9	60.5	91.2	67	JL1
Intensity		343372	75329	14355	13112	11333	10835	JL1
3.259	cyclohexane derivative	68.8	110.5	40.8	54.8	39.9	109.5	JL2
Intensity		200828	98286	56486	43800	21650	20192	JL2
3.293	No Hits	68.8	110.5	40.7	54.9	72.9	109.5	JL1
Intensity		187022	93645	51954	45164	38777	33595	JL1
3.295	cyclohexane derivative	82.8	81.5	40.8	55	68.8	67.1	JL2
Intensity		57543	54604	44200	37752	27266	16509	JL2
3.385	No Hits	68.8	110.6	40.8	81	80	40	JL2
Intensity		65605	42256	38196	22157	20869	16820	JL2
3.388	No Hits	82.7	81.2	54.8	40.7	94.4	67.1	JL1
Intensity		71342	49425	46212	32104	24388	21198	JL1
3.401	No Hits	68.7	110.7	40.7	95.1	54.8	82.9	JL1
Intensity		79123	46302	32926	25437	17563	15417	JL1
3.694	cyclohexane derivative	68.8	110.6	40.9	54.9	55.7	69.8	JL2
Intensity		420667	263212	134452	75528	51784	45857	JL2
3.718	Cyclohexane based structure	68.8	110.5	40.8	54.9	69.5	55.8	JL1
Intensity		557060	313287	154414	90250	58606	53022	JL1
3.853	o-xylene	91.1	105.9	51.1	95	92.1	50.1	JL1, JL2
Intensity		178467	31359	22155	20926	18778	15835	JL1, JL2
3.892	No Hits	91	42.9	40.9	68.9	39.9	56.8	JL2
Intensity		39824	38583	37137	17122	16658	11727	JL2
3.928	No Hits	42.9	91.1	40.9	56.7	39.9	68.9	JL2
Intensity		40776	34587	33285	21944	16551	14095	JL2

4.037	p-xylene	91.1	106	105.2	77	79.1	63.1	JL1, JL2
Intensity		986701	673788	203207	125484	89969	89466	JL1, JL2
4.208	cyclohexane derivative	68.7	110.6	40.8	54.8	109.6	55.8	JL1, JL2
Intensity		242073	140196	91611	70897	48670	47192	JL1, JL2
4.298	cyclohexane derivative	68.8	110.6	40.9	55	69.7	55.8	JL1, JL2
Intensity		234336	150607	90181	61144	52457	45460	JL1, JL2
4.38	cyclohexane derivative	68.8	40.9	110.6	54.9	96.8	55.6	JL1, JL2
Intensity		134391	62426	61584	41977	31677	30650	JL1, JL2
4.403	No Hits	96.7	54.9	40.9	68.7	95	81.1	JL1, JL2
Intensity		386988	159513	88429	69592	61835	55288	
4.56	No Hits	91.1	106	104.1	103.1	51	78.1	JL2
Intensity		496872	282183	273617	114632	109333	104686	JL2
4.584	No Hits	91.1	96.7	95.9	95.2	54.9	106	JL1
Intensity		123852	118046	81973	73509	61148	55015	JL1
4.819	methylated alkane	56.8	40.8	42.8	84.3	56	69.6	JL1, JL2
Intensity		219390	210265	167512	110838	77514	69293	JL1, JL2
4.918	long-chain alcohol	40.9	95.9	68.8	69.7	95	54.9	JL1, JL2
Intensity		80773	70373	64078	60966	58243	52783	JL1, JL2
4.99	ring structure hex/pent?	96.7	54.9	96	40.8	68.8	81.1	JL1, JL2
Intensity		354015	143703	138952	119654	110705	52697	JL1, JL2
5.065	No Hits	96.7	40.8	54.7	95.6	68.8	81.3	JL2
Intensity		71914	44296	42097	37619	25441	16411	JL2
5.315	hydrocarbon	81.1	67.1	95	40.8	79.1	54.8	JL1, JL2
Intensity		138357	101463	50289	39826	32399	22631	JL1, JL2
5.292	No Hits	81.1	67.1	40.8	95.1	79.2	82.7	JL2
Intensity		114456	76152	41499	32179	26439	25109	JL2
5.626	No Hits	81.3	82.7	54.8	40.9	67	80.1	JL2
Intensity		252663	232464	150312	123647	91454	43895	JL2
5.641	hydrocarbon	81.4	82.8	54.8	67.1	40.9	42.7	JL1
Intensity		258208	219386	170555	107408	89729	29573	JL1
5.743	No Hits	56.7	40.8	70.2	42.7	55.8	69	JL2
Intensity		76799	67353	53884	436110	43079	29600	JL2
5.767	methylated alkane	70.6	40.9	56.8	69.8	42.7	56.1	JL2
Intensity		84270	81475	67379	51185	50615	43765	JL2
5.888	methylated alkane	68.7	40.8	56.7	69.8	110.7	54.8	JL1, JL2
Intensity		94801	57368	51100	46479	38399	31745	JL1, JL2
5.951	No Hits	68.8	40.8	69.6	55	40	123.9	JL2
Intensity		37730	34090	15989	14523	13962	13904	JL2
6.101	methylated alkane	67.1	81.2	95	40.9	68.8	91.1	JL1
Intensity		37957	36486	31509	26964	19809	16199	JL1

6.259	No Hits	68.8	40.8	109	40	81	54.9	JL2
Intensity		35764	26351	18319	17004	13661	13551	JL2
6.444	No Hits	68.8	40.9	69.8	124.1	82.6	54.9	JL1
Intensity		102491	79972	63519	57792	56967	55572	JL1
6.548	No Hits	105.1	68.7	40.8	110.5	42.7	39.8	JL2
Intensity		78147	56681	50850	35728	26467	23242	JL2
6.62	No Hits	105.1	77	68.8	51.1	50.1	40.9	JL1, JL2
Intensity		369808	127277	101328	84515	71609	53668	JL1, JL2
6.759	No Hits	40.8	56.8	81.1	42.8	94.1	69.7	JL1, JL2
Intensity		70699	59519	59219	35049	30223	30092	JL1, JL2
6.81	No Hits	105.2	40.8	39.9	120.1	77	56.8	JL2
Intensity		37096	26752	18061	15052	11954	11811	JL2
6.955	No Hits	81.1	40.8	95.6	54.9	56.7	96.6	JL1, JL2
Intensity		85814	82037	57815	56125	54329	53652	JL1, JL2
7.057	No Hits	68.8	95.1	40.8	93.1	81	43	JL1, JL2
Intensity		62912	39001	35089	28456	28160	27256	JL1, JL2
7.355	No Hits	96.6	95.9	54.9	68.9	40.9	81	JL1
Intensity		159496	88711	87586	65683	57867	44802	JL1
7.649	1,2,3-trimethylbenzene	105.1	120.1	40.9	78.9	91.1	81.1	JL2
Intensity		87025	40835	20708	19599	19591	17172	JL2
7.748	No Hits	96.6	68.8	40.9	54.9	69.8	81.1	JL2
Intensity		62380	45061	34919	28566	20692	18409	JL2
7.993	methylated alkane	56.7	40.8	42.6	70.6	55.8	69.6	JL1, JL2
Intensity		244921	224585	171377	115562	87022	84462	JL1, JL2
8.146	No Hits	40.8	68.8	81	67.1	42.8	55.7	JL1, JL2
Intensity		112893	82726	67777	61728	59046	55369	JL1, JL2
8.78	No Hits	40.8	119.1	68.9	70.6	82.6	81.2	JL1, JL2
Intensity		43589	29976	26175	24792	24648	23696	JL1, JL2
8.952	Limonene	93	67.1	79.1	93.9	67.9	91.1	JL1, JL2
Intensity		256339	183602	121910	115308	113682	85016	JL1, JL2
9.096	methylated alkene	40.8	56.8	82.7	69.8	54.8	55.9	JL2
Intensity		269060	233240	159255	145344	138924	119930	JL2
9.124	No Hits	56.7	40.8	82.7	69.7	54.9	55.8	JL1
Intensity		170952	162019	147143	96162	84410	71105	JL1
9.25	Benzyl Alcohol	79	107	108	77.1	91	51	JL2
Intensity		365225	147843	145629	108074	64417	46995	JL2
9.412	No Hits	40.9	78.9	69.3	54.8	55.8	39.8	JL2
Intensity		28730	13418	13289	12641	12585	12151	JL2
9.939	cyclohexane derivative	81	95.1	67.1	137.5	40.9	81.8	JL1, JL2
Intensity		53469	43313	41422	28893	24595	22413	JL1, JL2
10.038	No Hits	40.8	119.2	42.8	56.8	54.9	40	JL2
Intensity		32811	19861	13861	13150	12623	11208	JL2

10.22	No Hits	40.9	70	42.7	39.8	91.2	54.9	JL2
Intensity		22795	13971	12043	12010	10354	10318	JL2
10.283	No Hits	40.8	81.1	105.2	68.9	54.8	91.1	JL2
Intensity		27329	14759	15640	13112	12305	12253	JL2
10.374	2-chloro-acetophenone	105.1	77.1	51.1	95.1	106.1	50.1	JL1, JL2
Intensity		229381	89340	51799	14471	14130	12957	JL1, JL2
10.682	No Hits	40.9	68.9	56.8	39.8	91.1	67	JL2
Intensity		37139	14832	14186	13278	13012	11882	JL2
10.818	No Hits	40.9	68.9	81.1	58.9	56	43	JL1, JL2
Intensity		31806	24958	20556	16968	16250	15479	JL1, JL2
11.245	No Hits	43	121	91	40.8	40	77	JL2
Intensity		40442	23593	14095	13205	12044	11346	JL2
11.545	No Hits	57.9	43	40.8	56.6	39.9	68.8	JL2
Intensity		39401	34853	27887	13695	13186	9835	JL2
11.891	methylated Alkane	40.8	56.8	42.8	70.5	69.8	68.8	JL1, JL2
Intensity		71550	47530	36284	31987	27841	16323	JL1, JL2
12.117	3-nonen-1-ol	40.9	81	95	69.6	68.7	67.2	JL1, JL2
Intensity		235463	160904	154245	144443	113063	112285	JL1, JL2
12.653	No Hits	40.9	82	39.8	91.1	68.9	79.2	JL2
Intensity		14242	13378	9817	5914	5862	5771	JL2
14.971	No Hits	128.2	81.1	40.9	95	67.2	68.8	JL2
Intensity		42683	37222	29641	27860	21420	17054	JL2
16.337	No Hits	81	40.9	82.7	67.1	95	67.9	JL1, JL2
Intensity		218626	199072	174301	128542	108009	96154	JL1, JL2
19.797	2-methylnaphthalene	141.3	142.2	115.2	81.1	40.8	39.8	JL2
Intensity		65327	50861	15975	10199	10117	9889	JL2
20.571	methylated alkane	40.9	42.6	56.8	70.6	69.7	83.7	JL2
Intensity		43598	36564	35702	31053	15271	14384	JL2
20.598	methylated alkane	56.8	40.9	42.7	70.6	69.9	84.5	JL2
Intensity		45326	38008	31033	26728	19453	16188	JL2
22.413	No Hits	42.7	82.8	40.9	70.4	69.2	88.9	JL1, JL2
Intensity		96902	92545	76988	41989	37459	34836	JL1, JL2
23.368	No Hits	93.9	105.2	91	119.2	93.1	107.1	JL2
Intensity		45970	34948	32052	24639	23591	22704	JL2
23.631	No Hits	88.8	42.8	40.9	70.1	55.8	68.8	JL1, JL2
Intensity		103350	99539	86780	75326	71340	39594	JL1, JL2
23.643	No Hits	40.8	88.8	42.7	70.3	55.8	68.8	JL1
Intensity		88160	84659	79915	46974	46501	26877	JL1
24.988	No Hits	161.2	91.2	79	93.1	105.2	133.1	JL1
Intensity		72989	42819	29905	29126	28524	27920	JL1
27.418	No Hits	43	107.1	40.9	68.8	93.1	79.1	JL1, JL2

Intensity		65600	57528	44351	44051	31443	25708	JL1, JL2
30.01	No Hits	40.9	56.9	70.6	42.8	84.4	39.9	JL1
Intensity		29122	24595	23659	20017	19184	11434	JL1
30.847	No Hits	102.2	40.8	121.1	173.9	103.1	91.2	JL1
Intensity		50195	15511	10265	10032	9290	7424	JL1
34.613	No Hits	56.9	40.9	70.5	42.8	84.2	69	JL1, JL2
Intensity		29535	22613	21327	14579	11505	9736	JL1, JL2
35.14	Benzophenone	105.1	182.2	77.1	51.1	181.3	50.1	JL1
Intensity		104985	67135	56582	42688	34883	13854	JL1
35.43	No Hits	173.2	110.8	82.8	55	155.1	99	JL1
Intensity		317669	86004	73003	72804	50985	44676	JL1

Table 3 Compounds found in the breath of participant CV-16

	Compound				**M/Z**			**Observed in**
1.359	Benzene	78.4	43.1	41.2	51.4	50.4	77.3	RA1, RA2
Intensity		167682	80205	75927	61182	55601	50478	RA1, RA2
1.413	No Hits	41.1	43.2	55.9	40.2	69.9	42.1	RA1
Intensity		68451	45102	38185	23808	20769	16188	RA1
1.566	No Hits	77.3	73.1	74.1	45.3	56.1	57	RA1
Intensity		152462	142383	119111	76360	75779	62934	RA1
1.675	No Hits	45.2	43.3	88.9	41.2	77.3	42.4	RA1
Intensity		223741	199863	173564	24454	23078	22765	RA1
1.781	No Hits	43.2	45.2	55.2	73.3	83	41.2	RA1, RA2
Intensity		65559	54428	42507	28555	25778	23060	RA1, RA2
1.928	No Hits	40.3	41.2	43.3	55.1	69.5	45.2	RA1
Intensity		17707	16002	14648	10108	8684	8553	RA1
2.016	Methylated Alkane	45.3	43.3	44.1	73.3	40.2	42.2	RA2
Intensity		119508	74213	24773	23217	21513	14024	RA2
2.052	No Hits	45.3	73.2	43.3	41.3	40.3	67.3	RA2
Intensity		90346	53775	42877	33801	29258	18276	RA2
2.092	No Hits	40.3	43.3	73.2	41.2	44.3	91.3	RA1
Intensity		20815	18734	18122	13337	8822	6296	RA1
2.223	Toluene	91.4	92.3	63.4	65.3	62.4	89.3	RA1, RA2
Intensity		677180	433847	107300	77830	43887	43412	RA1, RA2
2.431	No Hits	60.1	73.3	41.1	42.3	43.2	45.3	RA1, RA2
Intensity		64522	54809	31008	28678	24572	19705	
2.485	No Hits	151.4	149.3	45.1	73.2	133.4	43.3	RA2
Intensity		131887	120765	110944	70647	51549	45616	RA2
2.683	Hexanal	41.2	44.1	55.9	43.3	71.7	57	RA1, RA2
Intensity		115768	93759	76929	40271	33689	29085	
3.317	No Hits	80.3	81.3	40.3	73.3	50.3	45.4	RA2
Intensity		67990	43569	24569	15611	13128	9599	
3.743	o-xylene	91.4	106.2	51.4	105.4	40.3	69.4	RA1, RA2

Intensity		111122	27550	19411	15544	13456	11919	RA1, RA2
3.943	p-xylene	91.4	106.2	105.4	77.3	103.4	79.5	RA1, RA2
Intensity		166965	89719	50127	24691	18443	17061	RA1, RA2
4.098	No Hits	102.4	40.2	43.4	73.2	91.6	50.5	RA1
Intensity		23020	14459	9399	8701	7716	6648	
4.457	No Hits	91.5	106.3	105.4	104.4	45.3	51.4	RA1, RA2
Intensity		134068	118495	61236	43994	38290	30726	RA1, RA2
4.575	No Hits	42.2	55.2	69	80.2	41.1	83.2	RA1, RA2
Intensity		58192	48782	41449	35444	34081	27704	RA1, RA2
4.618	No Hits	40.3	41.1	69.1	55.2	42.1	43.2	RA1
Intensity		13840	12374	9500	7402	7400	7309	
4.719	No Hits	72	81.9	43.3	83.4	41.2	56.9	RA1, RA2
Intensity		69724	69310	62619	39450	32061	20549	RA1, RA2
4.782	No Hits	151.4	44.2	133.5	41.2	156.7	69.5	RA1, RA2
Intensity		180946	49986	48212	47621	46917	39108	RA1, RA2
4.863	No Hits	57	41.1	45.3	56	69.4	40.4	RA2
Intensity		71915	48794	34117	26340	24845	22037	RA2
4.891	No Hits	41.2	57.1	40.3	81.3	56.4	45.2	RA1, RA2
Intensity		25030	17856	12827	9856	9670	9203	RA1, RA2
5.128	No Hits	60.2	48.5	40.3	151.5	59.1	43.2	RA1, RA2
Intensity		33718	13477	12185	10143	8897	5991	RA1, RA2
5.54	No Hits	93.3	91.4	77.3	92.3	79.4	121.4	RA1, RA2
Intensity		159830	79559	53401	49165	46103	27073	RA1, RA2
5.888	No Hits	41.1	45.3	56.8	40.4	86.8	56	RA1, RA2
Intensity		20023	15817	13191	11008	10319	6570	RA1, RA2
6.172	No Hits	91.4	40.2	73.1	41.1	69.2	43.3	RA1
Intensity		15872	12685	7260	5537	4530	4396	RA1
6.41	No Hits	105.4	40.3	41.1	91.4	84.3	83.3	RA1, RA2
Intensity		29121	18470	16872	10082	9969	9407	RA1, RA2
6.482	2-chloro-acetophenone	105.4	77.4	51.3	50.3	78.3	106.3	RA1, RA2
Intensity		1043236	299777	221100	136591	75560	54373	RA1, RA2
6.792	No Hits	93.2	41.2	79.3	91.3	40.2	77.4	RA2
Intensity		86125	39225	38488	37583	25677	22004	RA2
6.801	No Hits	91.4	93.3	41	40.2	77.5	71	RA1
Intensity		22256	20919	19685	19409	14637	10942	RA1
6.91	No Hits	93.3	41.1	79.4	69	91.4	121.3	RA1, RA2
Intensity		93621	43385	29989	25231	24874	23371	RA1, RA2
7.155	Phenol	94.3	66.4	65.4	62.4	63.4	51.4	RA1, RA2
Intensity		452523	137881	81757	42279	42193	34080	RA1, RA2
7.236	No Hits	103.3	76.3	94.4	60.2	41	55.2	RA1, RA2
Intensity		77405	24076	18155	16553	15171	12695	RA1, RA2
7.835	methylated alkane	57	41.1	43.1	70.9	69.9	56	RA1, RA2
Intensity		101972	99890	47307	45052	40633	34758	RA1, RA2
7.998	No Hits	41.2	69.1	81.3	55.9	57.1	67.4	RA1, RA2
Intensity		75195	51170	50998	48362	45255	45103	RA1, RA2

8.218	No Hits	146.5	148.4	111.4	149.4	75.4	73.3	RA2
Intensity		59014	27676	26955	18618	18170	16471	RA2
8.628	No Hits	119.4	41.2	91.4	71.1	57	42.9	RA1, RA2
Intensity		105840	27499	22047	18549	16186	15107	RA1, RA2
8.889	No Hits	93.2	81.3	95.4	111.2	43.3	139.2	RA1
Intensity		546703	256940	205239	191667	177075	176090	RA1
9.101	Benzyl Alcohol	79.3	107.2	108.2	77.4	91.4	51.4	RA1, RA2
Intensity		778113	395168	375799	301143	169092	160781	RA1, RA2
9.861	No Hits	41.1	56.9	43.1	70.4	119.4	84.1	RA1
Intensity		40201	20914	19839	16468	14868	12956	
9.904	No Hits	93.3	91.4	41	92.5	121.3	79.4	RA2
Intensity		53582	40045	27304	27117	26889	26808	
10.058	No Hits	41.1	70.1	40.3	43	42.2	69.2	RA1, RA2
Intensity		68413	58863	34869	33124	32119	31238	RA1, RA2
10.211	2-chloro-acetophenone	105.3	77.3	51.3	50.4	106.3	95.4	RA1, RA2
Intensity		508393	180397	97145	48787	47465	37072	RA1, RA2
10.501	No Hits	41.1	57	70.7	69.9	69	81.5	RA2
Intensity		48100	45721	27571	22748	19342	17284	RA2
10.627	No Hits	55.1	41.1	82.7	68.9	56.1	70.1	RA2
Intensity		84179	59058	45655	40651	38578	37357	RA2
10.643	No Hits	41.2	83.1	55.2	69.1	56.1	119.4	RA1
Intensity		28897	23552	18361	18096	14927	14787	RA1
11.06	No Hits	43.3	121.3	51.4	77.3	105.3	50.4	RA1, RA2
Intensity		93717	61762	43982	40010	22192	20463	RA1, RA2
11.171	No Hits	132.4	117.5	91.4	115.5	40.3	105.3	RA1
Intensity		43248	32562	16896	15958	9702	8811	
11.702	Methylated Alcohol	41	57.1	43.1	71	70	83.9	RA1, RA2
Intensity		149946	128221	103834	90628	65338	53610	RA1, RA2
11.928	No Hits	41.2	57.1	81.3	56.1	69	95.3	RA1, RA2
Intensity		411506	343323	315509	295921	274926	274240	RA1, RA2
12.262	No Hits	119.5	41.1	40.3	69.3	120.5	91.4	RA1, RA2
Intensity		99919	27630	16354	16201	14707	13828	RA1, RA2
12.641	No Hits	73.2	87.2	88.2	40.3	43.1	41.1	RA2
Intensity		107520	57964	22298	19325	19134	17890	RA2
14.783	No Hits	95.3	81.4	71.2	41.1	82.2	67.4	RA1, RA2
Intensity		900051	868788	353468	345704	295341	254346	RA1, RA2
14.981	No Hits	73.3	41.2	60.1	95.4	105.3	55	RA2
Intensity		44321	43510	41052	22590	20152	19777	RA2
15.857	2-ethyl-1-hexanol	57	41.1	43	70.4	55.8	84.3	RA1, RA2
Intensity		135166	125500	85880	71356	46303	41683	RA1, RA2
16.128	No Hits	41.1	81.9	55.1	67.3	83	95.4	RA1, RA2
Intensity		452382	402663	278819	256199	244959	236906	RA1, RA2
16.39	No Hits	41.1	56.8	69.9	43	68.7	40.2	RA2
Intensity		59396	37661	32375	25268	18748	16889	RA2
16.489	No Hits	94.3	138.1	95.2	77.4	66.3	41.2	RA2

Intensity		209706	105105	65379	57509	46632	32507	RA2
16.552	No Hits	135.5	69.3	94.4	136.4	137.4	40.3	RA2
Intensity		206287	56232	47340	30396	30092	25171	RA2
17.566	No Hits	82.1	93.3	41.2	73.3	81.4	40.3	RA2
Intensity		50474	33296	26542	21257	19982	19820	RA2
19.086	Methylated Alkane	41.1	70.7	43	57	69.9	84.4	RA2
Intensity		92824	92681	79533	73611	48408	48195	RA2
19.555	5-methyl-quinoline	141.5	142.5	115.5	139.5	143.5	75.3	RA1, RA2
Intensity		478993	425855	140653	49365	41754	39350	RA1, RA2
20.206	No Hits	141.5	142.6	115.5	69.1	40.3	70.2	RA2
Intensity		149043	129517	44541	30225	26942	21314	RA2
20.323	Methylated Alkane	57	41.2	70.8	43	70.1	84.5	RA2
Intensity		162248	123927	106150	86515	58996	46103	RA2
20.374	No Hits	104.3	76.3	57.1	70.3	43.1	50.5	RA1
Intensity		101772	28671	19155	18047	16575	16009	RA1
22.134	No Hits	83	42.9	41.2	69.7	70.8	56	RA1
Intensity		215689	114222	74839	51081	44030	43688	RA1
21.244	Methylated Alkane	57	41.2	70.6	42.9	84.7	56.2	RA2
Intensity		55092	53794	39481	38065	34416	27336	RA2
22.148	No Hits	83	43	41.2	70.4	55.9	69	RA2
Intensity		562367	287236	237081	163623	150926	135802	RA2
23.079	No Hits	105.4	94.2	133.4	91.4	119.5	107.3	RA2
Intensity		89114	84024	83368	79969	75192	66832	RA2
23.367	No Hits	56	89	43	41.1	70.7	73.2	RA1, RA2
Intensity		381667	329238	301476	262738	211881	134619	RA1, RA2
24.577	No Hits	156.6	141.6	155.5	93.4	69	91.3	RA2
Intensity		106471	105281	52131	34369	33246	30657	RA2
24.702	No Hits	161.5	105.3	93.3	91.3	107.4	119.5	RA2
Intensity		259685	184383	147051	137724	125338	113185	RA2
25.18	3-ethyl-quinoline	141.6	156.6	155.5	115.4	128.4	142.5	RA2
Intensity		264001	180990	64159	58821	38210	33296	RA2
25.369	No Hits	156.5	141.5	155.4	157.7	41.2	91.3	RA2
Intensity		128866	122792	46724	26763	24802	22116	RA2
27.14	No Hits	69.1	41.2	107.3	43.3	67.3	93.3	RA1, RA2
Intensity		128336	119089	113552	101241	55101	48643	RA1, RA2
29.71	Methylated Alkane	41.1	56.8	70.9	43	84.5	70	RA1, RA2
Intensity		84855	82719	81063	54019	50932	45876	RA1, RA2
29.728	No Hits	57.1	41.2	70.8	43	40.2	84.3	RA2
Intensity		109004	92782	86014	62611	45026	44756	RA2
34.29	Methylated Alkane	41.2	57	71	43	84.6	70	RA2
Intensity		99120	88914	64675	64021	32339	32200	RA2
34.857	No Hits	105.4	182.4	77.4	51.4	181.5	50.4	RA1
Intensity		90887	58721	36862	29704	19477	10697	RA1

Table 4 Compounds found in the breath of participant CV-17

	Compound	M/Z						Observed in
1.27	No Hits	79.3	80.3	77.3	41.2	67.3	50.3	KK1
Intensity		253586	145899	67618	60507	50527	47436	KK1
1.387	3,4-dimethyl-1-hexene	41	55.9	56.9	42.9	42.1	83.1	KK1
Intensity		1373344	107955	1022205	824792	605590	592496	KK1
1.431	methylated alkane	69.9	56.9	41	42.9	56	70.7	KK1
Intensity		1577545	1273441	1184770	1153572	814900	811904	KK1
1.494	methylated alkene	69.9	41	42.8	55	56	68.9	KK1
Intensity		464957	276879	208032	192053	187425	146212	KK1
1.521	No Hits	77.2	55.9	69.9	41.1	55.1	69.1	KK1
Intensity		568793	293009	289184	275076	194275	186828	KK1
1.584	methylated alkane	41	57	42.9	70.2	55.9	42	KK1
Intensity		1090641	995691	970485	925615	557512	422619	KK1
1.683	No Hits	88.8	45.1	43.2	89.8	41.2	42.1	KK1
Intensity		440460	367199	223747	59508	53243	34586	KK1
1.8	methylated alkane	82.8	55	41.1	81.3	56	69.9	KK1
Intensity		185380	144828	116760	67287	56137	49170	KK1
1.89	No Hits	88	45.2	73.1	69.2	41	89	KK1
Intensity		377468	311388	189930	114548	96942	78144	KK1
1.935	No Hits	79.3	77.3	41	43.2	69.9	93.9	KK1
Intensity		55039	44200	42223	25032	22631	20261	
1.971	No Hits	45.2	59	43.2	75	44.2	41.1	KK2
Intensity		402929	343738	148143	134595	101905	78643	
2.223	Toluene	91.3	92.3	63.3	65.3	93.4	50.4	KK1,KK2
Intensity		1278693	834978	175314	146742	116700	86115	KK1,KK2
2.65	No Hits	41	43	84.5	95.3	57	40.2	KK1,KK2
Intensity		61106	40101	23720	20170	17666	17609	KK1,KK2
2.686	hexanal	41.1	55.8	44.1	43.3	57	67.3	KK1,KK2
Intensity		90925	66033	56541	49190	41007	36987	KK1,KK2
3.706	No Hits	67.3	41.1	81.9	40.3	43.3	69.2	KK1,KK2
Intensity		71272	59417	35296	21855	17649	16406	KK1,KK2
3.733	o-xylene	91.3	106.2	105.6	67.4	41.2	77.3	KK1,KK2
Intensity		149688	41975	31094	28207	26285	17772	KK1,KK2
3.803	No Hits	43.3	41.1	91.4	69.4	93.3	40.3	KK1
Intensity		51480	31552	30248	24860	18777	15620	KK1,KK2
3.942	p-xylene	91.3	106.2	105.3	51.3	103.4	92.4	KK1,KK2
Intensity		238625	128158	82450	33208	25170	24581	KK1,KK2
4.032	No Hits	41.1	56	43	55.2	42	69.1	KK1,KK2
Intensity		47106	28693	27308	16115	15093	12823	KK1,KK2

4.105	No Hits	102.4	40.2	41.1	73.2	43.2	51.4	KK1,KK2
Intensity		28778	11126	8574	7567	7447	6846	KK1,KK2
4.263	No Hits	43.3	81.3	41.1	45.2	69.1	40.2	KK1
Intensity		37803	30005	26921	20542	19431	17006	
4.315	No Hits	43.2	81.2	79.4	41.2	40.2	109.3	KK2
Intensity		30428	20419	16174	15652	13497	8329	
4.371	No Hits	43.2	56.9	73.3	41.1	45.2	40.3	KK1
Intensity		33756	29905	20629	17593	16970	13328	
4.46	No Hits	91.4	104.3	106.2	41.1	103.4	78.3	KK1,KK2
Intensity		87277	53364	49941	44546	39400	32404	KK1,KK2
4.587	No Hits	81.3	41.2	55.1	69.8	42	79.3	KK1,KK2
Intensity		316098	86984	80328	56391	49511	33956	KK1,KK2
4.687	methylated alkane	41.1	57	43	69.3	40.2	84.6	KK1,KK2
Intensity		51778	43339	25282	21891	19032	16949	KK1,KK2
4.714	No Hits	43.2	72.1	81.8	41.1	83.6	56.8	KK2
Intensity		63354	44906	39349	31729	28490	24511	KK2
4.786	No Hits	41.1	69.8	43	44.2	54.8	81.2	KK2
Intensity		69434	57806	36150	35727	35656	31862	KK2
4.795	No Hits	41.1	69.9	81.4	55.1	43.2	40.2	KK1
Intensity		63743	37568	34642	32015	24542	24161	KK1
4.886	No Hits	56.9	41.1	40.3	82.1	81.3	45.1	KK1,KK2
Intensity		35390	22189	16808	15056	13088	11530	KK1,KK2
5.16	No Hits	93.2	41.1	91.4	67.4	40.1	79.3	KK1,KK2
Intensity		37600	24193	17205	16507	12866	12422	KK1,KK2
5.365	No Hits	93.3	91.4	92.4	77.4	79.3	94.4	KK1
Intensity		479952	256913	188136	110931	85918	66559	KK1
5.572	No Hits	93.3	135.2	81.4	91.4	121.4	136	KK1
Intensity		20817738	7847619	7131190	3844762	3710714	3617143	KK1
6.031	No Hits	93.3	121.1	107.3	79.4	94.4	91.3	KK1,KK2
Intensity		549522	283649	176787	175878	133037	124753	KK1,KK2
6.093	No Hits	81.3	96	41.1	71.1	57.1	70.2	KK1,KK2
Intensity		110288	75805	66057	43033	31922	30751	KK1,KK2
6.21	No Hits	69.2	41.1	94.2	972	91.4	55.1	KK1
Intensity		135807	131657	75062	54094	51799	51117	
6.418	No Hits	105.4	41	43.3	55.1	83.2	69.2	KK1,KK2
Intensity		125735	56303	39302	34085	27876	24012	KK1,KK2
6.508	2-chloro-acetophenone	105.3	77.3	51.4	50.3	106	74.3	KK1,KK2
Intensity		773336	237527	188190	97011	96876	56858	KK1,KK2
6.701	No Hits	105.4	40.2	120.3	91.3	119.4	41.1	KK2
Intensity		39199	15745	12372	8722	8636	7804	
7.21	No Hits	95.3	81.3	67.3	41.2	93.2	54.9	KK1,KK2
Intensity		98013	82806	68556	56572	44448	40508	
7.99	No Hits	41.1	81.9	43	56	67.2	95.3	KK1,KK2

Intensity		145317	83380	80780	68487	66074	60629	
8.343	No Hits	121.3	93.1	91.4	79.3	136.2	122.4	KK1,KK2
Intensity		189372	156293	107414	58621	52678	43198	
8.442	No Hits	105.3	59.2	119.4	79.4	91.4	73.2	KK1
Intensity		61965	34558	33282	26608	25769	22702	
8.634	1-methyl-3-(1-methylethyl)-benzene	119.4	91.4	117.3	120.5	140.7	63.3	KK1,KK2
Intensity		651901	134176	74152	61814	55161	48163	
9.187	Benzyl Alcohol	79.3	108.2	93.3	107.2	91.4	77.3	KK1,KK2
Intensity		766773	502977	486066	426052	350001	316912	KK1,KK2
9.565	No Hits	93.3	80.3	79.4	91.4	105.3	77.3	KK1,KK2
Intensity		222290	128861	115447	97289	66157	56833	KK1,KK2
9.925	No Hits	93.3	121.3	77.3	91.4	135.8	92.3	KK1,KK2
Intensity		356225	161290	109860	103800	70295	57100	KK1,KK2
10.061	No Hits	70	41.2	83.1	55	42.9	82.1	KK1
Intensity		66400	65482	45262	35273	33952	33192	KK1
10.232	No Hits	105.3	77.4	51.3	106.4	50.4	120.5	KK1
Intensity		550210	134638	69771	52581	48124	46162	KK1
10.439	No Hits	93.2	79.4	93.9	91.4	121.2	81.3	KK1
Intensity		156528	92507	71774	65962	60750	55370	
10.668	Octanol?	41.1	68.9	82.9	55.1	56.1	69.8	KK2
Intensity		85830	69947	61385	59358	51102	46673	
10.709	methylated alkane	41.1	56.1	69.1	55.1	69.8	82.8	KK1
Intensity		199065	124848	117741	105368	99947	91318	
10.961	No Hits	121.4	93.3	79.3	91.4	136.2	77.5	KK1,KK2
Intensity		254627	219781	105301	96968	92549	58267	
11.094	No Hits	43.2	121.2	93.3	94.1	79.4	51.4	KK2
Intensity		52310	29869	25162	21042	18254	17374	
11.177	No Hits	117.5	132.4	115.5	43.2	91.4	121.3	KK1,KK2
Intensity		294459	197609	144943	99965	98229	72375	KK1,KK2
11.799	No Hits	93.2	41	80.3	71.2	121.3	55.1	KK1,KK2
Intensity		512393	282326	200794	190843	155585	124176	KK1,KK2
11.943	No Hits	41.1	81.3	70	57	56	95.2	KK1,KK2
Intensity		398453	336968	276864	267341	216246	190056	KK1,KK2
15.643	No Hits	81.4	95.3	41.2	67.4	82.3	123.2	KK1
Intensity		785862	547121	450446	348858	281348	240374	KK1
15.75	No Hits	93.2	121.5	136.1	81.5	95.4	59.2	KK1
Intensity		1008598	674669	468121	364815	329101	277579	KK1
16.217	methylated alcohol	81.4	41	95.3	83.2	67.4	43.1	KK1,KK2
Intensity		595069	422703	370705	335543	317941	261862	KK1,KK2
16.397	No Hits	81.4	43.2	126.1	41.1	169.7	71.1	KK1,KK2
Intensity		553676	472926	408267	229797	178913	141362	KK1,KK2

16.412	methylated alkane	41	82.9	69	43.2	57.1	55.1	KK1,KK2
Intensity		99846	74899	69027	64789	50030	49831	
16.468	No Hits	41.2	70.1	83	43.3	68.8	55.9	KK1
Intensity		181351	141874	132076	131403	128660	113873	
16.63	2-phenoxy-ethanol	94.4	138.2	95.2	66.4	51.3	77.3	KK1,KK2
Intensity		595976	215404	185352	169894	153552	134795	
17.386	No Hits	153.3	81.3	152.3	109.4	137.1	67.3	KK1
Intensity		2353570	1712552	1649672	1392099	1206058	1036254	KK1
17.484	No Hits	81.9	67.3	82.9	41.1	57	84.4	KK1
Intensity		1983667	1437238	856714	709762	625714	427000	KK1
17.601	No Hits	82	93.2	41.1	107.3	54	91.4	KK2
Intensity		156617	86242	67694	67212	58386	48542	
17.655	No Hits	82	93.3	107.4	79.5	106.4	54.2	KK1
Intensity		329160	193283	136391	136390	113149	106948	KK1
17.799	No Hits	41.1	56.1	84.9	56.9	82.9	103.2	KK1
Intensity		80112	56744	51713	45709	41996	39422	KK1
18.059	No Hits	110.3	82.1	95.3	153.2	137.5	109.3	KK1
Intensity		2428741	1611139	1448978	1253995	1003400	876273	KK1
18.93	No Hits	81.3	95.3	67.3	41.1	43.1	123	KK2
Intensity		81185	75222	46273	41409	40578	34762	KK2
19.14	No Hits	41.1	57.1	70.9	70	73.2	84.2	KK1
Intensity		125526	113215	92367	78987	63390	61845	KK1
19.599	No Hits	142.6	141.5	115.6	63.4	143.6	41	KK1
Intensity		700266	650877	168667	87700	57459	53539	KK1
19.796	No Hits	95.3	81.4	137.7	83	123.2	96.2	KK1
Intensity		5810952	4831429	2629643	2319881	2305714	1259048	KK1
20.255	No Hits	141.5	142.6	41	115.4	83.1	43	KK2
Intensity		225189	205512	58080	47695	34727	32900	KK2
20.354	No Hits	43	57	41.2	95.3	70.8	81.2	KK1,KK2
Intensity		243315	199414	186532	181933	181508	113473	KK1,KK2
20.524	No Hits	93.3	121.4	135.8	107.3	94.2	79.4	KK1,KK2
Intensity		101086	77591	67525	55445	49132	44519	KK1,KK2
21.262	methylated alkane	57.1	41.2	43.1	70.7	70	56	KK1
Intensity		89576	70759	68017	60875	51673	41133	KK1
22.181	No Hits	83.1	43.1	41.1	88.9	70.7	70	KK1,KK2
Intensity		763093	415957	410552	226542	221426	159716	
23.1	No Hits	119.5	94.1	91.3	105.4	107.3	95.2	KK1
Intensity		165243	163296	98639	95910	93528	93405	KK1
23.397	No Hits	89.2	43	56	41.2	68.9	70.7	KK1
Intensity		472747	418595	416375	376832	288996	274000	KK1
23.693	No Hits	79.4	80.2	123.4	81.4	161.3	41.2	KK1
Intensity		150571	129269	89877	86401	66553	53352	KK1

24.603	No Hits	156.5	141.5	155.5	95.2	79.4	161.3	KK1,KK2
Intensity		229401	197566	110750	65721	57090	47985	
25.205	7-ethyl-quinoline	141.5	156.5	155.6	153.5	115.4	128.6	KK1,KK2
Intensity		368888	309986	121740	69060	53551	48735	KK1,KK2
27.159	No Hits	43.3	107.3	93.3	41.2	69.2	95.4	KK1,KK2
Intensity		237623	204044	194110	146202	143227	95465	KK1,KK2
29.74	methylated alkane	41.1	70.9	57.1	43	84.7	70.1	KK1,KK2
Intensity		184029	155381	147124	123352	104086	91267	KK1,KK2
34.307	No Hits	57	41.2	70.9	84.6	43	56.1	KK1,KK2
Intensity		174338	125401	105985	102678	84231	69217	KK1,KK2
34.867	benzophenone	105.3	182.4	77.3	51.4	181.5	183.2	KK1,KK2
Intensity		247643	119727	113589	63777	51284	21112	KK1,KK2

Table 5 Compounds found in the breath of all participants included in the preliminary evaluation

	Compounds	**M/Z**						**Observed in**
1.27	No Hits	79.3	80.3	77.3	41.2	67.3	50.3	CV-17
1.359	Benzene	78.4	43.1	41.2	51.4	50.4	77.3	CV-14, CV-08, CV-16
1.387	3,4-dimethyl-1-hexene	41	55.9	56.9	42.9	42.1	83.1	CV-17
1.413	No Hits	41.1	43.2	55.9	40.2	69.9	42.1	CV-16
1.431	methylated alkane	69.9	56.9	41	42.9	56	70.7	CV-17
1.494	methylated alkene	69.9	41	42.8	55	56	68.9	CV-17
1.521	No Hits	77.2	55.9	69.9	41.1	55.1	69.1	CV-17
1.548	No Hits	40.9	43	73	74	54.9	69.7	CV-14
1.566	No Hits	77.3	73.1	74.1	45.3	56.1	57	CV-16
1.584	methylated alkane	41	57	42.9	70.2	55.9	42	CV-17
1.608	No Hits	72.8	73.8	54.7	56.7	55.6	74.7	CV-08
1.611	No Hits	40.8	42.9	77.1	39.9	69.7	70.5	CV-14
1.675	No Hits	45.2	43.3	88.9	41.2	77.3	42.4	CV-16
1.683	No Hits	88.8	45.1	43.2	89.8	41.2	42.1	CV-17
1.683	No Hits	76.9	44.9	72.8	77.8	69.3	56.5	CV-08
1.728	No Hits	44.9	43	88.7	40.8	42	60.9	CV-14
1.739	No Hits	93.9	92.3	44.9	96.4	91.2	94.9	CV-08
1.781	No Hits	43.2	45.2	55.2	73.3	83	41.2	CV-16
1.8	methylated alkane	82.8	55	41.1	81.3	56	69.9	CV-17, CV-14
1.826	1-ethyl-1-methyl-cyclopentane	82.7	54.9	40.9	81.6	68.9	69.8	CV-13

1.89	No Hits	88	45.2	73.1	69.2	41	89	CV-17
1.926	No Hits	82.4	54.6	80.8	66.9	93.7	79.7	CV-08
1.927	No Hits	45	39.9	40.9	87.9	43	73	CV-14
1.928	No Hits	40.3	41.2	43.3	55.1	69.5	45.2	CV-16
1.935	No Hits	79.3	77.3	41	43.2	69.9	93.9	CV-16, CV-17
1.971	No Hits	45.2	59	43.2	75	44.2	41.1	CV-16, CV-17
1.972	No Hits	43.1	40.9	69.7	40.1	54.9	73.1	CV-13
1.982	No Hits	79.1	40.9	42.9	95	39.9	77.1	CV-14
2.016	Methylated Alkane	45.3	43.3	44.1	73.3	40.2	42.2	CV-16
2.052	No Hits	45.3	73.2	43.3	41.3	40.3	67.3	CV-16
2.054	No Hits	43.1	40.9	40	69.7	72.9	55	CV-13
2.092	No Hits	40.3	43.3	73.2	41.2	44.3	91.3	CV-16
2.272	Toluene	91.1	92.1	63.1	65	50.1	51.1	CV-14, CV-08, CV-16, CV-17, CV-13
2.429	No Hits	43.1	60	96.7	73	40	40.9	CV-13
2.431	No Hits	60.1	73.3	41.1	42.3	43.2	45.3	CV-16
2.434	No Hits	96.6	40.9	54.8	40	68.9	91.2	CV-14
2.485	No Hits	151.4	149.3	45.1	73.2	133.4	43.3	CV-16
2.65	No Hits	41	43	84.5	95.3	57	40.2	CV-17
2.695	No Hits	43	40.9	40	42	84.5	55.8	CV-13
2.706	No Hits	40.8	42.6	96.7	84.2	55	41.8	CV-14
2.741	No Hits	40.9	43	43.9	55.9	40	81.8	CV-13
2.742	Hexanal	40.8	55.8	42.9	43.9	71.6	54.8	CV-14, CV-08, CV-16, CV-17
2.75	No Hits	93.9	175	176.9	176	93	91.8	CV-08
2.824	1,1-dimethyl-cyclohexane	96.6	40.8	54.9	39.9	43	55.7	CV-14
2.841	No Hits	96.1	42.9	95.1	54.9	40.9	91.1	CV-14
2.877	No Hits	94.3	95	43	93.7	96.3	44.9	CV-14
2.958	No Hits	43	42	94.9	60.5	91.2	67	CV-14
3.003	No Hits	94.4	96.7	44.9	95.6	92.6	76.8	CV-08
3.259	cyclohexane derivative	68.8	110.5	40.8	54.8	39.9	109.5	CV-14
3.293	No Hits	68.8	110.5	40.7	54.9	72.9	109.5	CV-14
3.295	cyclohexane derivative	82.8	81.5	40.8	55	68.8	67.1	CV-14
3.317	No Hits	80.3	81.3	40.3	73.3	50.3	45.4	CV-16
3.385	No Hits	68.8	110.6	40.8	81	80	40	CV-14
3.388	No Hits	82.7	81.2	54.8	40.7	94.4	67.1	CV-14

3.401	No Hits	68.7	110.7	40.7	95.1	54.8	82.9	CV-14
3.434	No Hits	94.4	72.8	73.7	44.9	54.8	96.7	CV-08
3.694	cyclohexane derivative	68.8	110.6	40.9	54.9	55.7	69.8	CV-14
3.718	Cyclohexane based structure	68.8	110.5	40.8	54.9	69.5	55.8	CV-14
3.706	No Hits	67.3	41.1	81.9	40.3	43.3	69.2	CV-17
3.853	o-xylene	91.1	105.9	51.1	95	92.1	50.1	CV-14, CV-08, CV-16, CV-17, CV-13
3.892	No Hits	91	42.9	40.9	68.9	39.9	56.8	CV-14
3.803	No Hits	43.3	41.1	91.4	69.4	93.3	40.3	CV-14, CV-17
4.037	p-xylene	91.1	106	105.2	77	79.1	63.1	CV-14, CV-08, CV-16, CV-17, CV-13
4.032	No Hits	41.1	56	43	55.2	42	69.1	CV-17
4.105	No Hits	102.4	40.2	41.1	73.2	43.2	51.4	CV-16, CV-17
4.208	cyclohexane derivative	68.7	110.6	40.8	54.8	109.6	55.8	CV-14
4.263	No Hits	43.3	81.3	41.1	45.2	69.1	40.2	CV-17
4.298	cyclohexane derivative	68.8	110.6	40.9	55	69.7	55.8	CV-14
4.315	No Hits	43.2	81.2	79.4	41.2	40.2	109.3	CV-17
4.38	cyclohexane derivative	68.8	40.9	110.6	54.9	96.8	55.6	CV-14
4.371	No Hits	43.2	56.9	73.3	41.1	45.2	40.3	CV-17
4.403	No Hits	96.7	54.9	40.9	68.7	95	81.1	CV-14
4.422	No Hits	102	75.9	73.9	49.9	63	74.9	CV-08
4.457	No Hits	91.5	106.3	105.4	104.4	45.3	51.4	CV-14, CV-08, CV-16, CV-17, CV-13
4.575	No Hits	42.2	55.2	69	80.2	41.1	83.2	CV-16
4.584	No Hits	91.1	96.7	95.9	95.2	54.9	106	CV-14, CV-08
4.587	No Hits	81.3	41.2	55.1	69.8	42	79.3	CV-17
4.618	No Hits	40.3	41.1	69.1	55.2	42.1	43.2	CV-16
4.687	methylated alkane	41.1	57	43	69.3	40.2	84.6	CV-17
4.714	No Hits	43.2	72.1	81.8	41.1	83.6	56.8	CV-17
4.719	No Hits	72	81.9	43.3	83.4	41.2	56.9	CV-16
4.782	No Hits	151.4	44.2	133.5	41.2	156.7	69.5	CV-16
4.786	No Hits	41.1	69.8	43	44.2	54.8	81.2	CV-17
4.792	styrene	104	78	91	103	105.9	50.9	CV-08
4.795	No Hits	41.1	69.9	81.4	55.1	43.2	40.2	CV-17
4.819	methylated alkane	56.8	40.8	42.8	84.3	56	69.6	CV-14
4.851	No Hits	72.9	55.7	54.8	69.7	68.8	96.5	CV-08

4.863	No Hits	57	41.1	45.3	56	69.4	40.4	CV-16
4.878	No Hits	43	41	40	69.8	44	54.9	CV-13
4.886	No Hits	56.9	41.1	40.3	82.1	81.3	45.1	CV-16, CV-17
4.918	long-chain alcohol	40.9	95.9	68.8	69.7	95	54.9	CV-14
4.99	ring structure hex/pent?	96.7	54.9	96	40.8	68.8	81.1	CV-14
5.065	No Hits	96.7	40.8	54.7	95.6	68.8	81.3	CV-14
5.076	No Hits	56.7	55.8	84.5	54.8	73	69.8	CV-08
5.128	No Hits	60.2	48.5	40.3	151.5	59.1	43.2	CV-16
5.16	No Hits	93.2	41.1	91.4	67.4	40.1	79.3	CV-17
5.292	No Hits	81.1	67.1	40.8	95.1	79.2	82.7	CV-14
5.315	hydrocarbon	81.1	67.1	95	40.8	79.1	54.8	CV-14
5.365	No Hits	93.3	91.4	92.4	77.4	79.3	94.4	CV-16, CV-17
5.572	No Hits	93.3	135.2	81.4	91.4	121.4	136	CV-17
5.626	No Hits	81.3	82.7	54.8	40.9	67	80.1	CV-14
5.641	hydrocarbon	81.4	82.8	54.8	67.1	40.9	42.7	CV-14
5.654	No Hits	93.1	92.1	91.2	40	79.2	43	CV-13
5.743	No Hits	56.7	40.8	70.2	42.7	55.8	69	CV-14
5.767	methylated alkane	70.6	40.9	56.8	69.8	42.7	56.1	CV-14
5.888	No Hits	41.1	45.3	56.8	40.4	86.8	56	CV-16
5.888	methylated alkane	68.7	40.8	56.7	69.8	110.7	54.8	CV-14
5.951	No Hits	68.8	40.8	69.6	55	40	123.9	CV-14
5.962	No Hits	92.9	91	91.9	78.9	151	77	CV-08
6.031	No Hits	93.3	121.1	107.3	79.4	94.4	91.3	CV-17
6.093	No Hits	81.3	96	41.1	71.1	57.1	70.2	CV-17
6.101	methylated alkane	67.1	81.2	95	40.9	68.8	91.1	CV-14
6.172	No Hits	91.4	40.2	73.1	41.1	69.2	43.3	CV-16
6.21	No Hits	69.2	41.1	94.2	109	91.4	55.1	CV-14, CV-17
6.418	No Hits	105.4	41	43.3	55.1	83.2	69.2	CV-16, CV-17
6.444	No Hits	68.8	40.9	69.8	124.1	82.6	54.9	CV-14
6.482	2-chloro-acetophenone	105.4	77.4	51.3	50.3	78.3	106.3	CV-16, CV-17
6.548	No Hits	105.1	68.7	40.8	110.5	42.7	39.8	CV-14
6.597	Benzoyl Bromide	105.2	77.1	51.1	50.1	78.2	74.2	CV-13
6.608	No Hits	90.9	149	72.9	91.9	151	119.6	CV-08
6.62	No Hits	105.1	77	68.8	51.1	50.1	40.9	CV-14
6.701	No Hits	105.4	40.2	120.3	91.3	119.4	41.1	CV-17

6.759	No Hits	40.8	56.8	81.1	42.8	94.1	69.7	CV-14
6.792	No Hits	93.2	41.2	79.3	91.3	40.2	77.4	CV-16
6.801	No Hits	91.4	93.3	41	40.2	77.5	71	CV-16
6.81	No Hits	105.2	40.8	39.9	120.1	77	56.8	CV-14
6.9	2-chloro-acetophenone	104.9	76.9	50.9	94.9	49.9	105.7	CV-08
6.91	No Hits	93.3	41.1	79.4	69	91.4	121.3	CV-16
6.955	No Hits	81.1	40.8	95.6	54.9	56.7	96.6	CV-14
7.057	No Hits	68.8	95.1	40.8	93.1	81	43	CV-14
7.143	No Hits	94	105	120	66	95	63	CV-08
7.155	Phenol	94.3	66.4	65.4	62.4	63.4	51.4	CV-16
7.21	No Hits	95.3	81.3	67.3	41.2	93.2	54.9	CV-17
7.236	No Hits	103.3	76.3	94.4	60.2	41	55.2	CV-16
7.251	Phenol	94.2	66.2	65.1	63.2	40	50.1	CV-13
7.355	No Hits	96.6	95.9	54.9	68.9	40.9	81	CV-14
7.378	Phenol	94	66	92.9	64.9	91	63	CV-08
7.416	No Hits	94	66	105	92.9	64.9	91	CV-08
7.597	Phenol	93.9	66	64.9	62.9	61.9	94.9	CV-08
7.649	1,2,3-trimethylbenzene	105.1	120.1	40.9	78.9	91.1	81.1	CV-14
7.668	Benzonitrile	103	75.9	66	93.9	50	75	CV-08
7.748	No Hits	96.6	68.8	40.9	54.9	69.8	81.1	CV-14
7.835	methylated alkane	57	41.1	43.1	70.9	69.9	56	CV-16
7.99	No Hits	41.1	81.9	43	56	67.2	95.3	CV-16, CV-17
7.993	methylated alkane	56.7	40.8	42.6	70.6	55.8	69.6	CV-14
8.134	No Hits	41	42.8	81.4	55.8	67.2	41.9	CV-13
8.146	No Hits	40.8	68.8	81	67.1	42.8	55.7	CV-14
8.218	No Hits	146.5	148.4	111.4	149.4	75.4	73.3	CV-16
8.343	No Hits	121.3	93.1	91.4	79.3	136.2	122.4	CV-17
8.442	No Hits	105.3	59.2	119.4	79.4	91.4	73.2	CV-17
8.628	No Hits	119.4	41.2	91.4	71.1	57	42.9	CV-16
8.634	1-methyl-3-(1-methylethyl)-benzene	119.4	91.4	117.3	120.5	140.7	63.3	CV-17
8.78	No Hits	40.8	119.1	68.9	70.6	82.6	81.2	CV-14
8.889	No Hits	93.2	81.3	95.4	111.2	43.3	139.2	CV-16
8.952	Limonene	93	67.1	79.1	93.9	67.9	91.1	CV-14, CV-13
9.083	methylated alkane	56.8	40.9	69.8	82.7	55.8	54.9	CV-13
9.096	methylated alkene	40.8	56.8	82.7	69.8	54.8	55.9	CV-14

9.101	Benzyl Alcohol	79.3	107.2	108.2	77.4	91.4	51.4	CV-16, CV-17
9.124	No Hits	56.7	40.8	82.7	69.7	54.9	55.8	CV-14
9.229	Benzyl Alcohol	79.1	108	107.1	77.1	51.2	91.1	CV-13
9.249	No Hits	56.7	55.7	69.8	54.8	82.7	57.7	CV-08
9.25	Benzyl Alcohol	79	107	108	77.1	91	51	CV-14
9.332	Limonene	92.9	79	67	93.8	67.8	90.9	CV-08
9.412	No Hits	40.9	78.9	69.3	54.8	55.8	39.8	CV-14
9.509	Hydrocarbon	56.7	69.8	70.6	55.7	54.8	82.7	CV-08
9.565	No Hits	93.3	80.3	79.4	91.4	105.3	77.3	CV-17
9.646	Benzyl Alcohol	78.9	76.9	106.9	107.9	72.8	50.9	CV-08
9.861	No Hits	41.1	56.9	43.1	70.4	119.4	84.1	CV-16
9.925	No Hits	93.3	121.3	77.3	91.4	135.8	92.3	CV-16, CV-17
9.939	cyclohexane derivative	81	95.1	67.1	137.5	40.9	81.8	CV-14
10.038	No Hits	40.8	119.2	42.8	56.8	54.9	40	CV-08
10.058	No Hits	41.1	70.1	40.3	43	42.2	69.2	CV-16, CV-17
10.075	No Hits	56.8	69.8	70.6	55.8	54.8	82.7	CV-14
10.211	2-chloro-acetophenone	105.3	77.3	51.3	50.4	106.3	95.4	CV-16, CV-17
10.22	No Hits	40.9	70	42.7	39.8	91.2	54.9	CV-14
10.283	No Hits	40.8	81.1	105.2	68.9	54.8	91.1	CV-14
10.374	2-chloro-acetophenone	105.1	77.1	51.1	95.1	106.1	50.1	CV-14, CV-13
10.439	No Hits	93.2	79.4	93.9	91.4	121.2	81.3	CV-17
10.501	No Hits	41.1	57	70.7	69.9	69	81.5	CV-16
10.643	No Hits	41.2	83.1	55.2	69.1	56.1	119.4	CV-16
10.668	Octanol?	41.1	68.9	82.9	55.1	56.1	69.8	CV-16, CV-17, CV-14
10.709	methylated alkane	41.1	56.1	69.1	55.1	69.8	82.8	CV-17
10.746	2-chloro-acetophenone	104.9	76.9	50.9	94.9	49.9	105.9	CV-08
10.798	No Hits	40.9	68.9	55.8	40	82.7	43	CV-13
10.818	No Hits	40.9	68.9	81.1	58.9	56	43	CV-14
10.961	No Hits	121.4	93.3	79.3	91.4	136.2	77.5	CV-17
11.06	No Hits	43.3	121.3	51.4	77.3	105.3	50.4	CV-16, CV-17
11.094	No Hits	43.2	121.2	93.3	94.1	79.4	51.4	CV-17
11.171	No Hits	132.4	117.5	91.4	115.5	40.3	105.3	CV-16, CV-17
11.177	No Hits	117.5	132.4	115.5	43.2	91.4	121.3	CV-17
11.214	No Hits	43.1	121.1	40	77.2	51.2	73	CV-13

11.245	No Hits	43	121	91	40.8	40	77	CV-14
11.545	No Hits	57.9	43	40.8	56.6	39.9	68.8	CV-14
11.702	Methylated Alcohol	41	57.1	43.1	71	70	83.9	CV-14, CV-16
11.799	No Hits	93.2	41	80.3	71.2	121.3	55.1	CV-17
11.87	No Hits	40.9	42.7	56.9	70.6	68.8	69.6	CV-13
11.928	No Hits	41.2	57.1	81.3	56.1	69	95.3	CV-16
11.943	No Hits	41.1	81.3	70	57	56	95.2	CV-17
12.091	Nonanal	40.9	81.1	69.8	55.7	56.8	67.2	CV-13
12.117	3-nonen-1-ol	40.9	81	95	69.6	68.7	67.2	CV-14
12.262	No Hits	119.5	41.1	40.3	69.3	120.5	91.4	CV-16
12.52	No Hits	81.1	66.9	69.7	95	55.7	56.8	CV-08
12.641	No Hits	73.2	87.2	88.2	40.3	43.1	41.1	CV-16
12.653	No Hits	40.9	82	39.8	91.1	68.9	79.2	CV-14
14.783	No Hits	95.3	81.4	71.2	41.1	82.2	67.4	CV-16
14.971	No Hits	128.2	81.1	40.9	95	67.2	68.8	CV-14
14.981	No Hits	73.3	41.2	60.1	95.4	105.3	55	CV-16
15.643	No Hits	81.4	95.3	41.2	67.4	82.3	123.2	CV-17
15.75	No Hits	93.2	121.5	136.1	81.5	95.4	59.2	CV-17
15.857	2-ethyl-1-hexanol	57	41.1	43	70.4	55.8	84.3	CV-16
16.128	No Hits	41.1	81.9	55.1	67.3	83	95.4	CV-16
16.217	methylated alcohol	81.4	41	95.3	83.2	67.4	43.1	CV-17
16.311	hydrocarbon	81.2	40.9	82	54.9	82.8	95.1	CV-13
16.337	No Hits	81	40.9	82.7	67.1	95	67.9	CV-14
16.39	No Hits	41.1	56.8	69.9	43	68.7	40.2	CV-16
16.397	No Hits	81.4	43.2	126.1	41.1	169.7	71.1	CV-17
16.412	methylated alkane	41	82.9	69	43.2	57.1	55.1	CV-17
16.489	No Hits	94.3	138.1	95.2	77.4	66.3	41.2	CV-16
16.468	No Hits	41.2	70.1	83	43.3	68.8	55.9	CV-17
16.552	No Hits	135.5	69.3	94.4	136.4	137.4	40.3	CV-16
16.63	2-phenoxy-ethanol	94.4	138.2	95.2	66.4	51.3	77.3	CV-17
16.79	No Hits	80.9	67	67.9	82.8	54.8	81.9	CV-08
16.813	methylated alkane	80.9	67	67.9	54.7	82.8	81.8	CV-08
17.386	No Hits	153.3	81.3	152.3	109.4	137.1	67.3	CV-17
17.484	No Hits	81.9	67.3	82.9	41.1	57	84.4	CV-17
17.601	No Hits	82	93.2	41.1	107.3	54	91.4	CV-16, CV-17
17.655	No Hits	82	93.3	107.4	79.5	106.4	54.2	CV-17
17.799	No Hits	41.1	56.1	84.9	56.9	82.9	103.2	CV-17

18.059	No Hits	110.3	82.1	95.3	153.2	137.5	109.3	CV-17
18.93	No Hits	81.3	95.3	67.3	41.1	43.1	123	CV-17
19.086	Methylated Alkane	41.1	70.7	43	57	69.9	84.4	CV-16
19.14	No Hits	41.1	57.1	70.9	70	73.2	84.2	CV-17
19.555	5-methyl-quinoline	141.5	142.5	115.5	139.5	143.5	75.3	CV-16, CV-17
19.796	No Hits	95.3	81.4	137.7	83	123.2	96.2	CV-17
19.797	2-methylnaphthalene	141.3	142.2	115.2	81.1	40.8	39.8	CV-14
20.255	No Hits	141.5	142.6	41	115.4	83.1	43	CV-16, CV-17
20.296	No Hits	117.1	90	141.1	142.1	88.9	62.9	CV-08
20.323	Methylated Alkane	57	41.2	70.8	43	70.1	84.5	CV-16, CV-17
20.524	No Hits	93.3	121.4	135.8	107.3	94.2	79.4	CV-17
20.556	No Hits	104.1	40.9	56.9	42.8	76.1	70.5	CV-13
20.571	methylated alkane	40.9	42.6	56.8	70.6	69.7	83.7	CV-14
20.598	methylated alkane	56.8	40.9	42.7	70.6	69.9	84.5	CV-14
21.244	Methylated Alkane	57	41.2	70.6	42.9	84.7	56.2	CV-16
21.262	methylated alkane	57.1	41.2	43.1	70.7	70	56	CV-17
22.148	No Hits	83	43	41.2	70.4	55.9	69	CV-16
22.181	No Hits	83.1	43.1	41.1	88.9	70.7	70	CV-16, CV-17
22.385	No Hits	82.9	42.8	41	70.2	69.5	55.9	CV-13
22.413	No Hits	42.7	82.8	40.9	70.4	69.2	88.9	CV-14
23.079	No Hits	105.4	94.2	133.4	91.4	119.5	107.3	CV-16
23.1	No Hits	119.5	94.1	91.3	105.4	107.3	95.2	CV-17
23.368	No Hits	93.9	105.2	91	119.2	93.1	107.1	CV-14
23.397	No Hits	89.2	43	56	41.2	68.9	70.7	CV-16, CV-17
23.611	No Hits	40.9	88.9	42.8	55.7	70.4	73	CV-13
23.631	No Hits	88.8	42.8	40.9	70.1	55.8	68.8	CV-14
23.643	No Hits	40.8	88.8	42.7	70.3	55.8	68.8	CV-14
23.693	No Hits	79.4	80.2	123.4	81.4	161.3	41.2	CV-17
23.958	No Hits	91	119	105	92.9	72.9	93.9	CV-08
24.577	No Hits	156.6	141.6	155.5	93.4	69	91.3	CV-16, CV-17
24.702	No Hits	161.5	105.3	93.3	91.3	107.4	119.5	CV-16
24.988	No Hits	161.2	91.2	79	93.1	105.2	133.1	CV-14
25.205	7-ethyl-quinoline	141.5	156.5	155.6	153.5	115.4	128.6	CV-16, CV-17
25.369	No Hits	156.5	141.5	155.4	157.7	41.2	91.3	CV-16
27.14	No Hits	69.1	41.2	107.3	43.3	67.3	93.3	CV-16
27.159	No Hits	43.3	107.3	93.3	41.2	69.2	95.4	CV-17
27.418	No Hits	43	107.1	40.9	68.8	93.1	79.1	CV-14

28.027	No Hits	107	93	69	79	151	91	CV-08
29.74	methylated alkane	41.1	70.9	57.1	43	84.7	70.1	CV-16, CV-17
29.993	No Hits	56.9	41	70.7	42.8	40	84.1	CV-13
30.01	No Hits	40.9	56.9	70.6	42.8	84.4	39.9	CV-14
30.847	No Hits	102.2	40.8	121.1	173.9	103.1	91.2	CV-14
34.29	Methylated Alkane	41.2	57	71	43	84.6	70	CV-16, CV-17
34.613	No Hits	56.9	40.9	70.5	42.8	84.2	69	CV-14
34.857	benzophenone	105.4	182.4	77.4	51.4	181.5	50.4	CV-16, CV-17, CV-14
35.122	Benzophenone	105.2	182.2	77.1	181.4	51.2	40	CV-13
35.43	No Hits	173.2	110.8	82.8	55	155.1	99	CV-14

References

1. Nicola Silvestris, G.S., Anna Maria Sebastio, Carla Broussard, Salvatore Pisconti, Gianmauro Numico, Giuseppe Pezzella, *Lack of response to a cisplatin-adriamycin based chemotherapy regimen in a patient affected by adenocarcinoma of the lung with rhabdoid features.* Cancer Therapy, 2007. **5**(B): p. 377-378.
2. W.D. Travis, L.B.T., S.S. Devesa, *Lung Cancer.* Cancer, 1995. **75**(1 Suppl): p. 191-202.
3. Barbone, F., et al., *Cigarette Smoking and Histologic Type of Lung Cancer in Men.* Chest, 1997. **112**(6): p. 1474-1479.
4. Wikipedia. Available from: http://en.wikipedia.org/wiki/File:Lung_cancer.jpg.
5. Brambilla, E., et al., *The new World Health Organization classification of lung tumours.* European Respiratory Journal, 2001. **18**(6): p. 1059-1068.
6. Microscopyu.com. Available from: http://www.microscopyu.com/staticgallery/pathology/oatcellcarcinoma10x01_b.html.
7. M.C. Champaneria, I.M.M., M. Kidd, G.N. Eick, *Friedrich Feyrter: a precise intellect in a diffuse system.* Neuroendocrinology, 2006. **83**(5-6): p. 394-404.
8. gsi.ir. Available from: http://www.gsi.ir/Images/MedicalGeology/mesothelioma.jpg.
9. Williams, C., *Lung cancer: the facts*. 2nd ed. 1992: Oxford University press.
10. Canada, N.C.I.o., *Canadian cancer statistics 1996*, N.C.I.o. Canada, Editor. 1996: Toronto, ON.
11. U.K., C.R., *Cancer Stats Mortality - UK*, C.R. U.K., Editor. 2007.
12. Doll, R. and A.B. Hill, *Lung Cancer And Other Causes Of Death In Relation To Smoking: A Second Report On The Mortality Of British Doctors.* The British Medical Journal, 1956. **2**(5001): p. 1071-1081.
13. UK, C.R. *Smoking and cancer*. Available from: http://info.cancerresearchuk.org/healthyliving/smokingandtobacco/ (retrieved 16/12/2010).
14. U.K., C.R., *Cancer Stats - Key facts. Lung Cancer and Smoking*, C.R. U.K., Editor. 2009.
15. CANCER*Mondial*, *GLOBOCAN*.
16. Ng, N., et al., *Predicting lung cancer death in Africa and Asia: differences with WHO estimates.* Cancer Causes and Control, 2009. **20**(5): p. 721-730.
17. lungcancer.co.uk. Available from: http://www.lungcancer.co.uk/diagnosis.htm.
18. Ginsberg, M.S., et al., *Pulmonary Tumorlets: CT Findings.* Am. J. Roentgenol., 2004. **183**(2): p. 293-296.
19. patient.co.uk. Available from: http://www.patient.co.uk/health/Bronchoscopy.htm.
20. Campbell, N.A., *Biology*. 4th ed. 1996: The Benjamin/Cummings Publishing Company, Inc.
21. Miekisch, W., J.K. Schubert, and G.F.E. Noeldge-Schomburg, *Diagnostic potential of breath analysis--focus on volatile organic compounds.* Clinica Chimica Acta, 2004. **347**(1-2): p. 25-39.
22. Dreher, D. and A.F. Junod, *Role of oxygen free radicals in cancer development.* European Journal of Cancer, 1996. **32**(1): p. 30-38.
23. Knight, J., *Diseases related to oxygen-derived free radicals.* Ann Clin Lab Sci, 1995. **25**(2): p. 111-121.

24. Eiceman, G.A., et al., *Ion mobility spectrometry for continuous on-site monitoring of nicotine vapors in air during the manufacture of transdermal systems.* Journal of Hazardous Materials, 1995. **43**(1-2): p. 13-30.
25. Griffin, M.T. *Differential mobility spectroscopy for chemical agent detection*. in *Proceedings of SPIE - The International Society for Optical Engineering*. 2006.
26. Krebs, M.D., et al., *Detection of biological and chemical agents using Differential Mobility Spectrometry (DMS) technology.* IEEE Sensors Journal, 2005. **5**(4): p. 696-702.
27. Palmer, P.T. and T.F. Limero, *Mass spectrometry in the U.S. space program: past, present, and future.* Journal of the American Society for Mass Spectrometry, 2001. **12**(6): p. 656-675.
28. Petinarides, J., et al. *Implementation of a new technology for point detection*. in *Proceedings of SPIE - The International Society for Optical Engineering*. 2005.
29. Phillips, M., *Method for the collection and assay of volatile organic compounds in breath.* Analytical Biochemistry, 1997. **247**(2): p. 272-278.
30. Baumbach, J.I. and M. Westhoff, *Ion mobility spectrometry to detect lung cancer and airway infections.* Spectroscopy Europe, 2006. **18**(6): p. 22-27.
31. Chan, H.P., C. Lewis, and P.S. Thomas, *Exhaled breath analysis: Novel approach for early detection of lung cancer.* Lung Cancer, 2009. **63**(2): p. 164-168.
32. Di Natale, C., et al., *Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors.* Biosensors and Bioelectronics, 2003. **18**(10): p. 1209-1218.
33. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD.* Lung Cancer. **In Press, Corrected Proof**.
34. Peng, G., et al., *Diagnosing lung cancer in exhaled breath using gold nanoparticles.* Nat Nano, 2009. **4**(10): p. 669-673.
35. Phillips, M., et al., *Detection of lung cancer with volatile markers in the breath.* Chest, 2003. **123**(6): p. 2115-2123.
36. Phillips, M., et al., *Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study.* The Lancet, 1999. **353**(9168): p. 1930-1933.
37. Chen, X., et al., *A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis.* Cancer, 2007. **110**(4): p. 835-844.
38. SPONRING, A., et al., *Release of Volatile Organic Compounds from the Lung Cancer Cell Line NCI-H2087 In Vitro.* Anticancer Research, 2009. **29**(1): p. 419-426.
39. Barash, O., et al., *Sniffing the Unique "Odor Print" of Non-Small-Cell Lung Cancer with Gold Nanoparticles.* Small, 2009. **5**(22): p. 2618-2624.
40. Pauling, L., et al., *Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(10): p. 2374-2376.
41. Zlatkis, A., H. Lichtenstein, and A. Tishbee, *Concentration and analysis of trace volatile organics in gases and biological fluids with a new solid adsorbent.* Chromatographia, 1973. **6**(2): p. 67-70.
42. Gordon, S., et al., *Volatile organic compounds in exhaled air from patients with lung cancer.* Clin Chem, 1985. **31**(8): p. 1278-1282.
43. O'Neill, H., et al., *A computerized classification technique for screening for the presence of breath biomarkers in lung cancer.* Clin Chem, 1988. **34**(8): p. 1613-1618.

44. Phillips, M., *Detection of carbon disulfide in breath and air: a possible new risk factor for coronary artery disease.* International Archives of Occupational and Environmental Health, 1992. **64**(2): p. 119-123.
45. Å panÄ›l, P. and D. Smith, *Influence of water vapour on selected ion flow tube mass spectrometric analyses of trace gases in humid air and breath.* Rapid Communications in Mass Spectrometry, 2000. **14**(20): p. 1898-1906.
46. Basanta, M., et al., *Non-invasive metabolomic analysis of breath using differential mobility spectrometry in patients with chronic obstructive pulmonary disease and healthy smokers.* Analyst, 2010. **135**(2): p. 315-320.
47. Miekisch, W. and J.K. Schubert, *From highly sophisticated analytical techniques to life-saving diagnostics: Technical developments in breath analysis.* TrAC Trends in Analytical Chemistry, 2006. **25**(7): p. 665-673.
48. Cheng, W.-H. and W.-J. Lee, *Technology development in breath microanalysis for clinical diagnosis.* Journal of Laboratory and Clinical Medicine, 1999. **133**(3): p. 218-228.
49. Frank Kneepkens, C.M., G. Lepage, and C.C. Roy, *The potential of the hydrocarbon breath test as a measure of lipid peroxidation.* Free Radical Biology and Medicine, 1994. **17**(2): p. 127-160.
50. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD.* Lung Cancer, 2009. **64**(2): p. 166-170.
51. Baumbach, J.I., et al., *Metabolites in human breath: Ion mobility spectrometers as diagnostic tools for lung diseases*. Breath Analysis: for Clinical Diagnosis and ATherapeutic Monitoring, ed. A. Amann and D. Smith. 2005, Singapore: World Scientific Publ Co Pte Ltd. 53-66.
52. Ruzsanyi, V., et al., *Detection of human metabolites using multi-capillary columns coupled to ion mobility spectrometers.* Journal of Chromatography A, 2005. **1084**(1-2): p. 145-151.
53. Vautz, W. and J.I. Baumbach, *Analysis of bio-processes using ion mobility spectrometry.* Engineering in Life Sciences, 2008. **8**(1): p. 19-25.
54. G.A.S. *BreathSpec*. Available from: http://www.gas-dortmund.de/IMS-Products/BreathSpec/1_370.html.
55. Kearney, N., Richardson, Alison., *Nursing patients with cancer: principles and practice*. 2006: Elsevier.
56. Fauci, A.S., Braunwald, E., Kasper, D.L., Hauser, S.L., Longo, D.L., Jameson, J. L., Loscalzo, J., *Harrison's Principles of Internal Medicine*. 17 ed. 2008: McGraw-Hill Medical.
57. Anstie, F.E., *Final experiments on the elimination of alcohol from the body.* Practitioner, 1874. **13**(15).
58. Bogen, E., *Drunkenness: A quantitative study of acute alcoholic intoxication.* Journal of the American Medical Association, 1927. **89**.
59. Liljestrand, G., Linde, P., *Über die ausscheidung des alkohols mit der expirationsluft.* Scand. Arch. Physiol., 1930. **60**.
60. Harger, R.N., Lamb, E.B., Hulpieu, H.R., *A rapid chemical test for intoxication employing breath.* Journal of the American Medical Association, 1938. **110**.
61. Lucas, D.M., *Professor Robert F. Borkenstein - An Appreciation of his lif and work.* Forensic Science Reviews, 2000. **12**(1).

62. Riter, L.S., et al., *Direct analysis of volatile organic compounds in human breath using a miniaturized cylindrical ion trap mass spectrometer with a membrane inlet.* Rapid Communications in Mass Spectrometry, 2002. **16**(24): p. 2370-2373.
63. Blain, M.G., et al., *Towards the hand-held mass spectrometer: design considerations, simulation, and fabrication of micrometer-scaled cylindrical ion traps.* International Journal of Mass Spectrometry, 2004. **236**(1-3): p. 91-104.
64. Brinckerhoff, W.B., et al., *Miniature time-of-flight mass spectrometers for in situ composition studies.* Acta Astronautica. **52**(2-6): p. 397-404.
65. Chaudhary, A., et al., *Fabrication and testing of a miniature cylindrical ion trap mass spectrometer constructed from low temperature co-fired ceramics.* International Journal of Mass Spectrometry, 2006. **251**(1): p. 32-39.
66. Janfelt, C., R. Graesboll, and F.R. Lauritsen, *Characterization and optimization of membrane inlets for a miniature ion trap mass spectrometer operating at a high background pressure of humid air.* International Journal of Mass Spectrometry, 2008. **276**(1): p. 17-23.
67. Sillon, N. and R. Baptist, *Micromachined mass spectrometer.* Sensors and Actuators B: Chemical, 2002. **83**(1-3): p. 129-137.
68. Tadjimukhamedov, F.K., et al., *Evaluation of a Differential Mobility Spectrometer/Miniature Mass Spectrometer System.* Journal of the American Society for Mass Spectrometry, 2010. **21**(9): p. 1477-1481.
69. Detection, S. *IONSCAN500DT*. Available from: http://www.smithsdetection.com/IONSCAN_500DT.php (retrieved 11/12/2010).
70. Detection, S. *SABRE 4000*. Available from: http://www.smithsdetection.com/SABRE_4000.php (retrieved 11/12/2010).
71. Detection, S. *LCD-FR*. Available from: http://www.smithsdetection.com/LCD-FR.php.
72. Detection, S. *LCD Nexus*. Available from: http://www.smithsdetection.com/LCDNexus.php (retrieved 11/12/2010).
73. Detection, S. *CENTURION II*. Available from: http://www.smithsdetection.com/centurion_II.php (retrieved 11/12/2010).
74. Detection, S. *IONSCAN SENTINEL II*. Available from: http://www.smithsdetection.com/Sentinel.php (retrieved 11/12/2010).
75. Dynamics, G. *JUNO*. Available from: http://www.gdatp.com/Products/Detection_Systems/ChemBio/JUNO.asp (retrieved 11/12/2010).
76. Environics. *Environics*. Available from: http://www.environics.fi/index.php?page=company (retrieved 11/12/2010).
77. Zolotov, Y.A., *Ion mobility spectrometry.* Journal of Analytical Chemistry, 2006. **61**(6): p. 519.
78. O'Donnell, R.M., X. Sun, and P.d.B. Harrington, *Pharmaceutical applications of ion mobility spectrometry.* TrAC Trends in Analytical Chemistry, 2008. **27**(1): p. 44-53.
79. Wishart, D.S., et al., *HMDB: the Human Metabolome Database.* Nucleic Acids Research. **35**(suppl 1 http://nar.oxfordjournals.org/content/35/suppl_1/D521.abstract): p. D521-D526.
80. Smith, C.A., O'maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., *METLIN: A Metabolite Mass Spectral Database.* Therapeutic Drug Monitoring, 2005. **27**(6): p. 747-751.

81. Kanehisa, M., Goto, S., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Res., 2000. **28**: p. 27-30.
82. Eiceman, G.A., et al., *Pattern recognition analysis of differential mobility spectra with classification by chemical family.* Analytica Chimica Acta, 2006. **579**(1): p. 1-10.
83. Haldane, J.S. and J.G. Priestley, *The Regulation of the Lung-ventilation.* Journal of Physiology, 1905. **32**: p. 225-266.
84. Kischkel, S., et al., *Breath biomarkers for lung cancer detection and assessment of smoking related effects -- confounding variables, influence of normalization and statistical algorithms.* Clinica Chimica Acta, 2010. **411**(21-22): p. 1637-1644.
85. Leone, A.M., et al., *Nitric Oxide Is Present in Exhaled Breath in Humans: Direct GC-MS Confirmation.* Biochemical and Biophysical Research Communications, 1994. **201**(2): p. 883-887.
86. Kharitonov, S.A. and P.J. Barnes, *Biomarkers of some pulmonary diseases in exhaled breath.* Biomarkers, 2002. **7**(1): p. 1 - 32.
87. Phillips, M., et al., *Detection of lung cancer using weighted digital analysis of breath biomarkers.* Clinica Chimica Acta, 2008. **393**(2): p. 76-84.
88. Rizvi, N. and D.F. Hayes, *A "breathalyser" for lung cancer?* The Lancet, 1999. **353**(9168): p. 1897-1898.
89. Reynolds, J.C., et al., *Detection of Volatile Organic Compounds in Breath Using Thermal Desorption Electrospray Ionization-Ion Mobility-Mass Spectrometry.* Analytical Chemistry, 2010. **82**(5): p. 2139-2144.
90. Phillips, M., et al., *Increased breath biomarkers of oxidative stress in diabetes mellitus.* Clinica Chimica Acta, 2004. **344**(1-2): p. 189-194.
91. Phillips, M., et al., *Volatile biomarkers of pulmonary tuberculosis in the breath.* Tuberculosis, 2007. **87**(1): p. 44-52.
92. Phillips, M., et al., *Volatile markers of breast cancer in the breath.* Breast Diseases, 2004. **15**(1): p. 53.
93. Phillips, M. and J. Greenberg, *Method for the collection and analysis of volatile compounds in the breath.* Journal of Chromatography: Biomedical Applications, 1991. **564**(1): p. 242-249.
94. Robles, A.C., et al., *Analysis of Biomarker Features from a Differential Mobility Spectrometer for the Detection of Tuberculosis.* Life Science Systems and Applications Workshop, Bethesda, MO, 2007.
95. Wallace, L.A., et al., *Total exposure assessment methodology (team) study: Personal exposures, indoor-outdoor relationships, and breath levels of volatile organic compounds in New Jersey.* Environment International, 1986. **12**(1-4): p. 369-387.
96. Park, K.-H. and W.-K. Jo, *Personal volatile organic compound (VOC) exposure of children attending elementary schools adjacent to industrial complex.* Atmospheric Environment, 2004. **38**(9): p. 1303-1312.
97. Van Berkel, J.J.B.N., et al., *A profile of volatile organic compounds in breath discriminates COPD patients from controls.* Respiratory Medicine, 2010. **104**(4): p. 557-563.
98. Larstad, M.A.E., et al., *Determination of ethane, pentane and isoprene in exhaled air - effects of breath-holding, flow rate and purified air.* Acta Physiologica, 2007. **189**(1): p. 87-98.

99. Larstad, M., et al., *Determination of ethane, pentane and isoprene in exhaled air using a multi-bed adsorbent and end-cut gas-solid chromatography.* Analyst, 2002. **127**(11): p. 1440-1445.
100. Teshima, N., et al., *Determination of acetone in breath.* Analytica Chimica Acta, 2005. **535**(1-2): p. 189-199.
101. Ueta, I., et al., *Breath acetone analysis with miniaturized sample preparation device: In-needle preconcentration and subsequent determination by gas chromatography-mass spectroscopy.* Journal of Chromatography B, 2009. **877**(24): p. 2551-2556.
102. Tassopoulos, C.N., D. Barnett, and T. Russell Fraser, *BREATH-ACETONE AND BLOOD-SUGAR MEASUREMENTS IN DIABETES.* The Lancet, 1969. **293**(7609): p. 1282-1286.
103. van den Velde, S., et al., *Differences between alveolar air and mouth air.* Analytical Chemistry, 2007. **79**(9): p. 3425-3429.
104. Basanta, M., et al., *An adaptive breath sampler for use with human subjects with an impaired respiratory function.* Analyst, 2007. **132**(2): p. 153-163.
105. Anderson, J.C. and M.P. Hlastala, *Breath tests and airway gas exchange.* Pulmonary Pharmacology & Therapeutics, 2007. **20**(2): p. 112-117.
106. environsupply.com. Available from: http://www.environsupply.com/M10-18.htm (retrieved 31/10/2010).
107. McGarvey, L.J., Shorten, C.V., *The effects of adsorption on the reusability of Tedlar air sampling bags.* American Industrial Hygiene Association Journal, 2000. **61**(3): p. 375-380.
108. Beghi, S. and J.-M. Guillot, *Sample water removal method in volatile organic compound analysis based on diffusion through poly(vinyl fluoride) film.* Journal of Chromatography A, 2006. **1127**(1-2): p. 1-5.
109. Basanta, M., et al., *Increasing analytical space in gas chromatography-differential mobility spectrometry with dispersion field amplitude programming.* Journal of Chromatography A, 2007. **1173**(1-2): p. 129-138.
110. Thomson, J.J., *Rays of positive electricity and their application to chemical analysis.* Longmans Green, 1913. **London**.
111. Paul, W. and H. Steinwedel, **EIN NEUES MASSENSPEKTROMETER OHNE MAGNETFELD.* Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences, 1953. **8**(7): p. 448-450.
112. Stafford Jr, G.C., et al., *Recent improvements in and analytical applications of advanced ion trap technology.* International Journal of Mass Spectrometry and Ion Processes, 1984. **60**(1): p. 85-98.
113. Cooks, R.G., et al., *Mass spectrometers: instrumentation.* International Journal of Mass Spectrometry and Ion Processes, 1992. **118-119**: p. 1-36.
114. March, R.E., *An Introduction to Quadrupole Ion Trap Mass Spectrometry.* Journal of Mass Spectrometry, 1997. **32**(4): p. 351-369.
115. Miller, R.A., Nazarov, E. G., Zapata, A., Davis, C. E., Eiceman, G. A., Bashall, A. D., *Systems for differential ion mobility analysis*, U.S. Patent, Editor. Filed 2008, Sionex Corporation, The Charles Stark Draper Laboratory Inc.: United States.
116. Shvartsburg, A.A., *Differential ion mobility spectrometry*. 1 ed. 2009: CRC Press. 322.
117. Molina, M.A., et al., *Design-of-experiment optimization of exhaled breath condensate analysis using a miniature differential mobility spectrometer (DMS).* Analytica Chimica Acta, 2008. **628**(2): p. 155-161.

118. Cunha, M.G., Hoenigman, S., Kanchagar, C., Rearden, P., Sassetti, C. S., Trevejo, J. M., Keshava, N. *Joint analysis of differential mobility spectrometer and mass spectrometer features for tuberculosis biomarkers*. in *IEEE Eng Med Biol Soc*. 2008. Vancover, Canada.
119. Thomson, J.J., *Conduction of Electricity through Gases*. 1903, London: Cambridge University Press.
120. Albritton, D.L., et al., *Mobilities of Mass-Identified H3+ and H+ Ions in Hydrogen.* Physical Review, 1968. **171**(1): p. 94.
121. Siems, W.F., *Reconsidering the Mason-Schamp equation*, in *International Society for Ion Mobility Spectrometry*. 2010: Albuquerque, New Mexico.
122. Buryakov, I.A., et al., *A new method of separation of multi-atomic ions by mobility at atmospheric pressure using a high-frequency amplitude-asymmetric strong electric field.* International Journal of Mass Spectrometry and Ion Processes, 1993. **128**(3): p. 143-148.
123. Borsdorf, H., E.G. Nazarov, and G.A. Eiceman, *Atmospheric pressure chemical ionization studies of non-polar isomeric hydrocarbons using ion mobility spectrometry and mass spectrometry with different ionization techniques.* Journal of the American Society for Mass Spectrometry, 2002. **13**(9): p. 1078-1087.
124. McDaniel, E.A.M.a.E.W., *Transport Properties of ions in Gases*. 1988, New York: Wiley.
125. G.A. Eiceman, E.G.N., J.E. Rodriguez, J.F. Bergloff, *Positive Reactant Ion Chemistry for Analytical, High Temperature Ion Mobility Spectrometry (IMS): Effects of Electric Field of the Drift Tube and Moisture, Temperature, and Flow of the Drift Gas.* International Journal for Ion Mobility Spectrometry, 1998. **1**(1): p. 28-37.
126. Bocos-Bintintan, V., *Effect of water concentration on DMS responses*, G. Blackburn, Editor. 2010, Loughborough University Department of Chemistry.
127. Kim, S.H., K.R. Betty, and F.W. Karasek, *MOBILITY BEHAVIOR AND COMPOSITION OF HYDRATED POSITIVE REACTANT IONS IN PLASMA CHROMATOGRAPHY WITH NITROGEN CARRIER GAS.* Analytical Chemistry, 1978. **50**(14): p. 2006-2012.
128. Spangler, G.E. and C.I. Collins, *Reactant ions in negative ion plasma chromatography.* Analytical Chemistry, 1975. **47**(3): p. 393-402.
129. Song, G., et al., *Quantitative breath analysis of volatile organic compounds of lung cancer patients.* Lung Cancer, 2010. **67**(2): p. 227-231.
130. Pennazza, G., et al., *Interpretation of exhaled volatile organic compounds*, in *Exhaled Biomarkers*. 2010, European Respiratory Society Journals Ltd. p. 115-129.
131. Poli, D., et al., *Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study.* Respiratory Research, 2005. **6**(1): p. 71.
132. Chen, X., Cao, M., Li, Y., Hu, W., Wang, P., Ying, K., Pan, H., *A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method.* Meas. Sci. Technol., 2005. **16**(8).
133. Wehinger, A., et al., *Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas.* International Journal of Mass Spectrometry, 2007. **265**(1): p. 49-59.
134. Machado, R.F., et al., *Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath.* Am. J. Respir. Crit. Care Med., 2005. **171**(11): p. 1286-1291.

135. Diaries, D. *The Hippocratic Oath: Modern Version*. Available from: http://www.pbs.org/wgbh/nova/doctors/oath_modern.html (retrieved 15/12/2010).
136. Organisation, W.H. *Guidelines for contributors*. Available from: http://www.who.int/bulletin/volumes/84/current_guidelines/en/index1.html.
137. Herring, J., *Medical Law and Ethics*. 1 ed. 2006: OUP Oxford. 720.
138. Mason, J.K., McCall Smith, R. A., *Law and Medical Ethics*. 2 ed. 1987, London: Butterworths. 345.
139. Pattinson, S.D., *Medical Law and Ethics*. 2nd Revised edition ed. 2009: Sweet and Maxwell. 704.
140. Gantt, H.L., *Organizing for Work*. 1919, New York: Harcourt, Brace and Howe.
141. Jolliffe, I.T., *Principal Component Analysis*. second ed. Springer Series in Statistics. 2002: Springer-Verlag New York Inc.
142. Miller, J.a.M., J. C., *Statistics and Chemometrics for Analytical Chemistry*. 4th Edition ed. 2001: Prentice Hall. 288.
143. Amann, A., P. Å panÄ›l, and D. Smith, *Breath analysis: The approach towards clinical applications.* Mini-Reviews in Medicinal Chemistry, 2007. **7**(2): p. 115-129.
144. Amann, A., et al., *Applications of breath gas analysis in medicine.* International Journal of Mass Spectrometry, 2004. **239**(2-3): p. 227-233.
145. Cao, W. and Y. Duan, *Current status of methods and techniques for breath analysis.* Critical Reviews in Analytical Chemistry, 2007. **37**(1): p. 3-13.
146. Deng, C., X. Zhang, and N. Li, *Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography-mass spectrometry.* Journal of Chromatography B, 2004. **808**(2): p. 269-277.
147. Phillips, M., et al., *Prediction of lung cancer using volatile biomarkers in breath.* Cancer Biomarkers, 2007. **3**(2): p. 95-109.
148. Sanchez, J.M. and R.D. Sacks, *Development of a multibed sorption trap, comprehensive two-dimensional gas chromatography, and time-of-flight mass spectrometry system for the analysis of volatile organic compounds in human breath.* Analytical Chemistry, 2006. **78**(9): p. 3046-3054.
149. TOFWERKS. *IMS-TOF*. Available from: http://www.tofwerk.com/cms/index.php?Products:IMS-TOF (retrieved 12/12/2010).
150. Corporation, W. *SYNAPT G2 HDMS*. Available from: http://www.waters.com/waters/nav.htm?locale=en_GB&cid=10110057 (retrieved 12/12/2010).
151. Scientific, T. *FAIMS Interface*. Available from: http://www.thermoscientific.com/wps/portal/ts/products/detail?productId=11961722 (retrieved 12/12/2010).
152. Cheung, W.X., Y.; Thomas, C.L.; Goodacre, R., *Discrimination of bacteria using pyrolysis-gas chromatography-differential mobility spectrometry (Py-GC-DMS) and chemometrics.* Analyst - London - Society of Public Analysts then Royal Society of Chemistry, 2009. **134**(3): p. 557-563.
153. Prasad, S., et al., *Analysis of bacterial strains with pyrolysis-gas chromatography/differential mobility spectrometry.* Analyst, 2006. **131**(11): p. 1216-1225.

154. Zhao, W., et al., *Two-dimensional Wavelet Analysis Based Classification of Gas Chromatogram Differential Mobility Spectrometry Signals.* Analytica Chimica Acta. **In Press, Accepted Manuscript**.
155. Bell, S., et al., *Neural Network Recognition of Chemical Class Information in Mobility Spectra Obtained at High Temperatures.* Analytical Chemistry, 2000. **72**(6): p. 1192-1198.
156. Dynamics, N. *Progenesis LC-MS And advanced proteomics research solution*. Available from: http://www.nonlinear.com/products/progenesis/lc-ms/overview/ (retrieved 18/12/2010).
157. Sciences, G.L. *DeCyder™ MS Differential Analysis Software*. Available from: http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&parentid=976024&moduleid=166070&zone=Elpho (retrieved 18/12/2010).
158. Research, N.I.o.H., *Research in the NHS - Human Resources (HR) Good Practice Resource Pack.*
159. Office, H. *About CRB*. Available from: http://www.crb.homeoffice.gov.uk/about_crb.aspx (retrieved 14/12/2010).
160. Office, H. *Welcome to the Criminal Records Bureau*. Available from: http://www.crb.homeoffice.gov.uk/ (retrieved 14/12/2010).
161. Plebani, C., et al., *An optimized sampling and GC-MS analysis method for benzene in exhaled breath, as a biomarker for occupational exposure.* Talanta, 1999. **50**(2): p. 409-412.
162. Bairoch, A., Boeckmann, B., Ferro, S., Gasteiger, E., *Swiss-Prot: Juggling between evolution and stability.* Brief. Bioinform, 2004. **5**: p. 39-55.
163. Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., Merrill, A. H. Jr., Murphy, R. C., Raetz, C. R., Russell, D. W., Subramaniam, S., *LMSD: LIPID MAPS structure database.* Nucleic Acids Res., 2006. **35**: p. 527-32.
164. SpectralWorks. *AalyzerPro*. Available from: http://www.spectralworks.com/analyzerpro.asp (retrieved 04/01/11).
165. ACD/Labs. *ACD/IntelliXtract*. Available from: http://www.acdlabs.com/products/adh/ms/intellixtract/ (retrieved 11/10/10).
166. NIST. *AMDIS; Automated Mass-Spectral Deconvolution and Identification System*. Available from: http://www.amdis.net/ (retrieved 08/11/10).
167. National Institute of Standards and Technology. *National Institute of Standards and Technology*. Available from: http://www.nist.gov/index.html (retrieved 08/11/10).
168. National Institute for Health and Clinical Excellence. *Referral guidelines for suspected cancer*. Available from: http://www.nice.org.uk/nicemedia/live/10968/29814/29814.pdf (retrieved 08/11/10).
169. Lee, F.G., Morris, J. C., *Kinetics of Chlorination of phenol - chlorophenolic tastes and odors.* Int. J. Air Wat. Poll., 1962. **6**: p. 419-431.
170. Kováts, E., *Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone.* Helvetica Chimica Acta, 1958. **41**(7): p. 1915-1932.
171. UK, C.R. *Common cancers - UK mortality statistics*. Available from: http://info.cancerresearchuk.org/cancerstats/mortality/cancerdeaths/ (retrieved 16/12/2010).
172. UK, C.R., *Cancer Stats Mortality - UK*, C.R. UK, Editor. 2005.

173. UK, C.R. *Lung Cancer Survival statistics*. Available from: http://info.cancerresearchuk.org/cancerstats/types/lung/survival/ (retrieved 19/12/2010).
174. Griffin, J.L., et al., *Standard reporting requirements for biological samples in metabolomics experiments: Mammalian/in vivo experiments.* Metabolomics, 2007. **3**(3): p. 179-188.