Design and Analysis of a Biometric Access Control System using an Electronic Olfactory Device to Identify Human Odour Characteristics

Stephen McMillan

Thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the degree of Doctor of Philosophy

This research programme was carried out in collaboration with Mastiff Electronic Systems Ltd.

January 2000

Abstract

The use of an electronic olfactory device, termed an electronic 'nose', was investigated for the detection of unique human odour characteristics. The detection of these unique odours was applied to the field of biometrics for access control, where a human's unique characteristics were used to authenticate a user of an access control system. An electronic odour sensing device was designed and constructed using an array of conducting polymer gas sensors in order to facilitate the regular screening of a group of human subjects over a period of six weeks.

A static sampling method was used to measure odour levels from human hands, which were found to contain a reliable source of human odour. Human odour levels were low so dynamic sampling proved to be unsuitable for this application due to the dilution of the odour mixture. Feature analysis results revealed that the features of adsorption and desorption gradient contained discriminatory information in addition to the commonly used maximum divergence. Pattern recognition revealed that neural network architectures produced superior results when compared to statistical methods as a result of their ability to model the non-linearities in the data set. The highest recognition rate was 73% which was produced using a Multi-Layer Perceptron (MLP) neural network compared to 63% obtained using the best statistical method of Parzen windows. The majority of the recognition error was caused by a minority of the humans. Analysis of sensor data revealed that only 30% of the sensor array were contributing discriminatory information so it was deduced that performance would undoubtedly improve if a full array of effective sensors were available.

Exploratory data analysis revealed that human odour changed from day to day and often an increasing divergence with time was observed. A time-adaptive method was devised which increased the recognition to 89%, but was still too low for use as a biometric recognition device. However, use as a verification device demonstrated acceptable levels of performance but resulted in high levels of user frustration caused by a high proportion of users being falsely rejected. This work demonstrated that an olfactory based biometric access control system could be a realistic proposition but requires further work, especially in the areas of sensor development and unique human odour research, before an operational system could be produced.

Acknowledgements

I would like to thank Dr Richard Seals, University of Greenwich, for his help and guidance throughout the duration of my PhD.

I am grateful to Mastiff Electronic Systems Ltd. for giving me the opportunity to conduct PhD research into such an interesting area of work.

Finally, I would like to thank my wife, Judith, whose encouragement, help and support proved indispensable.

Contents

Nomenclature

Chapter 1. Introduction to Access Control

Symbol	Definition
U	Access control system User.
S	Access control system Subject.
P	Probability User, U, is subject S.
T	Biometric Template supplied by User, U, to represent Subject S.
O	Access control system Object, for example a door.
R	Access control system Rules applied to Object O.
Pr	Access control system privileges granted to Subject S.
FAR	False Acceptance Rate, performance measure for biometric verification devices. FAR reveals the percentage of access attempts which have been successfully authenticated when they should have been rejected recognised and is a function of acceptence threshold (Equation 1.11).
FRR	False Rejection Rate, performance measure for biometric verification devices. FRR reveals the percentage access attempts which have been rejected when they should have been successfully authenticated recognised and is a function of acceptence threshold (Equation 1.12).

Chapter 4. Data Analysis and Feature Extraction

Symbol	Definition
B	Sensor response baseline readings measured before the sensor response measurements, S, are taken.
S	Sensor response data measured from the Analogue to Digital converter on the data acquisition card. The range is $\pm$ 2096 corresponding to voltage readings of $\pm$ 3V.
B_m	Mean of the baseline measurements, B, from a given sensor.

R	Sensor response measured from the mean baseline reading, B_m, therefore eliminating the effect of sensor baseline drift.
F	Feature attribute calculated from the sensor response, R. There are 16 features numbered 0 to 15.
PCA	Principal Components Analysis: linear unsupervised feature extraction algorithm.
LDA	Linear Discriminant Analysis: linear supervised feature extraction algorithm.
SAM	Sammon Map: non-linear unsupervised feature extraction algorithm.
SOM	Kohonen Self Organising Map: non-linear unsupervised feature extraction algorithm.

Chapter 5. Pattern Recognition

Symbol	Definition
x	Input vector to pattern recognition system. In this case x is the optimised feature vector.
C	A class, person in this case, used in a pattern recognition system. There are c classes in total.
FAR	False Acceptance Rate (see previous definition).
FRR	False Rejection Rate (see previous definition).
RR	Recognition Rate, performance measure for biometric recognition devices. RR reveals the percentage of access control system users who have been correctly recognised and is a function of acceptence threshold (Equation 5.1).
ARR	Absolute Recognition Rate, performance measure for biometric recognition devices. RR is a single figure measure which represents the maximum percentage of access control system users who can be correctly recognised and is defined as the RR at an acceptence threshold of zero.
W	Weight vector used in statistical discriminant functions or Neural Computing methods.
a	Summed input to a neuron, of a neural network, in a hidden or output layer.

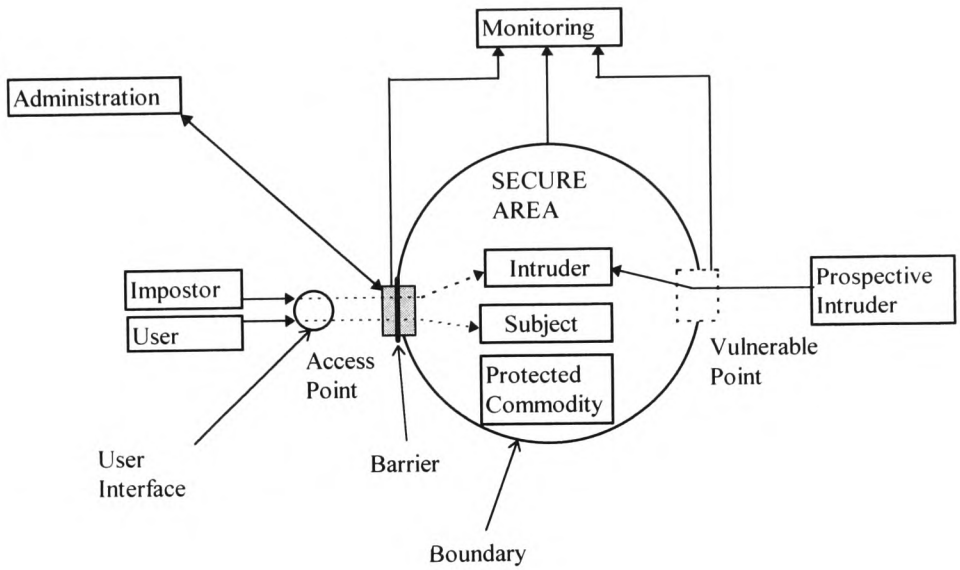
z	Output from a neuron, of a neural network, after applying a transfer function g upon the summed input a.
$\tilde{g}$	Neuron activation function at the output layer of a neural network.
g	Neuron activation function at the hidden layer of a neural network.
MLP	Multi-Layer Perceptron neural network.
LVQ	Learning Vector Quantisation neural network.
RBF	Radial Basis Function neural network.
50/10/40	Data set which is partitioned into 50% training data, 10% validation data and 40% test data.
10/5/85	Data set which is partitioned into 10% training data, 5% validation data and 85% test data.

1. Introduction

The function of access control is to prevent unauthorised entry but to allow authorised entry into a specifically defined area of space, the *secure area*. The term 'access' denotes the ability of a person to enter the secure area. The secure area is enclosed and defined by a *boundary* which serves to ensure that the area is not breached by unauthorised users, *impostors*. The boundary may be physical such as the walls of a computer room or could have 'virtual' boundaries which exist between the various areas of a computer program. The boundary usually possesses several *vulnerable points,* such as windows or uncontrolled doors, which require careful consideration to ensure that the secure area is not breached.

A system user gains access to the secure area via an *access point* comprising a *user interface* for the transfer of *authentication* data, which is used by the system to determine the identity of a user. A *barrier* is used to provide easy access into the secure area for authorised users but should be impenetrable when not activated by a successful access attempt. Examples of barriers include doors, turnstiles and software generated barriers which restrict access to data residing on a computer. If an unauthorised user gains entry into the secured area then the *protected commodity* becomes under risk. The protected commodity can be defined as any data or process which is for use by authorised personnel only. Examples of protected commodities include company pay role information residing on a database or physical money in a bank vault. Figure 1.1 illustrates the various constituents of an access control system.

Figure 1.1 An Access Control System



Guards have traditionally provided access control but are prone to error and can be coerced by criminals[1]. The lock and key was one of the first ventures into tag based access control but is not suitable for many access control requirements. For example, if a user loses a traditional key then the lock and *all* keys must be replaced; this process is not ideal especially when a large number of users are involved. However, electronic access control provides the ability to combat sophisticated impostors and can also provide a greater level of control than traditional methods. For example, if a user loses an electronic tag then the administrator can immediately deactivate the lost tag and re-issue a new tag to the single user concerned.

Most people will invariably encounter at least one form of electronic access control on a regular basis; for example, obtaining money from a cash machine and entering the workplace with an electronic tag. This work is applied to physical access control but most of the concepts are applicable to any form of access control, such as computer networks.

The *reliability* of an access control system measures the success of the system in preventing impostors gaining entry into a secure area via the access point, and consequently becoming intruders. The reliability of an access control system can be expressed in percentage form as:

$$Reliability = \left(1 - \frac{\sum acc_{\text{int}}}{\sum acc_{auth}}\right).100 \quad \% \qquad (1.1)$$

where

acc_{auth} is an authorised access into the secure area via the access control system.

acc_{int} is an unauthorised entry into the secure area via the access control system, called an intruder access.

Therefore, a perfect access control system would possess a reliability rating of 100%. An intruder access could be caused by numerous factors such as: unauthorised acquisition of a tag or password which grants the impostor access, production of a duplicate, 'fake' tag or attribute which grants the impostor access or entry to a secure area whilst an authorised user is currently passing through the secure boundary.

Access control provides a method of authorising access via designated access points but does not prevent intruders gaining entry into the secure area by other means. Other gateways into the secure area may be via one of the vulnerable points such as a window in a physical access control system or an uncontrolled access point. A measure of the integrity of a secure area is *security*, which is a function of the number of intruders entering a secure area by any means, not only via the access control system. Security of a controlled area can be described as :

$$Security = \left(1 - \frac{\sum entry_{\text{int}}}{\sum entry_{auth}}\right).100 \quad \% \qquad (1.2)$$

where

$entry_{auth}$ is any authorised access into the secure area.

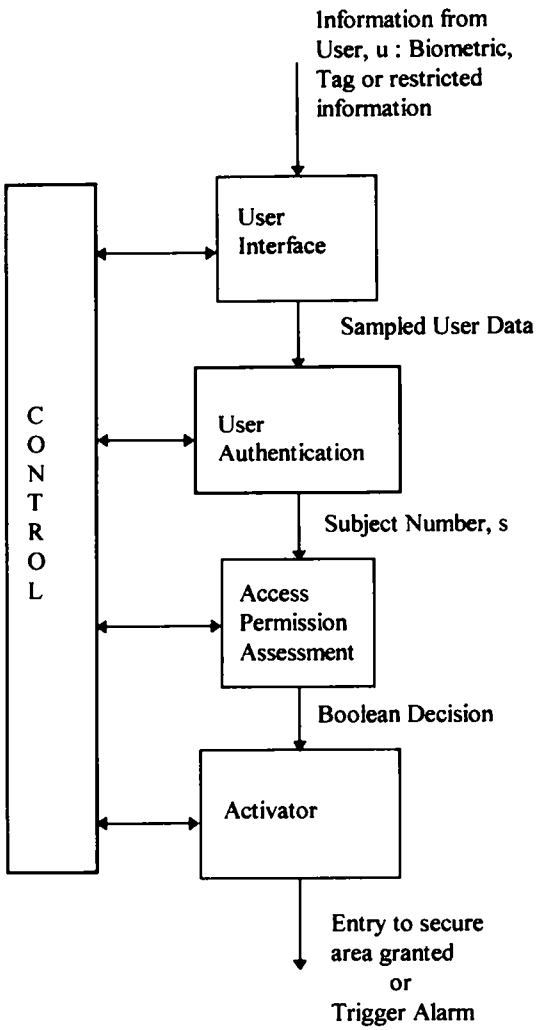
$entry_{int}$ is an unauthorised entry into the secure area by any means.

Therefore, a secure area which has never been compromised possesses a security rating of 100%. The security rating of a secure area can be improved by introducing any process or attribute to a secure area which reduces the probability of an intruder gaining access to the secure area. Access control improves security by reducing the probability of unauthorised entry through an access point, therefore access control be regarded as a subset of security. Security encompasses many disciplines each of which may contribute to the security rating of a secure area. Security features either act as a contribution to the integrity of the secure area which means making the area more impenetrable to intruders, for example electric fences and security glass in windows, or act as a deterrent which dissuades a potential intruder usually by increasing the probability of interception of the intruder, for example CCTV or Guard Dogs. Any secure area is only as safe as the weakest link in the system so each aspect of the secure environment must be analysed[2].

The concept of access control can also be extended to a role of monitoring vulnerable points such as windows in a physical access control system. The monitored points of the access control system provide a warning of possible unauthorised entry due a breach of the boundary, vulnerable point or an unauthorised breach to an access point. Monitoring can improve security by providing a deterrent to prospective intruders. For example, a prospective intruder may decide against breaching a boundary if he knows his entry will be reported immediately to a security guard. The use of electronic access control, to enhance security, is now becoming a necessity because of recent rises in crime in the workplace[3].

The access control process is composed of two distinct operations: *User Authentication* followed by *Access Permission Assessment*[4], as illustrated in figure 1.2.

Figure 1.2 The Access Control Process



User authentication determines the identity of the user before further access checks are made. The actual process of user authentication is dependent upon the method deployed at a particular access point but is generally concerned with confirming, or authenticating, a user's identity. User authentication will be discussed further in section 1.2. Once a user's identity is confirmed the user, u, is transformed into

subject, *s*, which is the user's identity within the access control system[4]. The access request of subject, *s*, is then assessed to determine whether the subject should be granted permission to access through the access point and into the controlled area. Access Permission Assessment will be discussed further in section 1.3.

There are many methods available to control access through the access point which may be as traditional as a lock and key[5] or as innovative as a retina scan technique[6]. The method chosen depends upon a variety of factors including cost and the consequences of unauthorised access to a restricted area[7]. However, it must be appreciated that the method of entry is only as secure as the associated entry point. For example, a sophisticated fingerprint verification system is useless if a door which it is controlling can be opened by the use of force.

Most access control systems currently use a form of personal tag issued to a user. The tags range greatly in complexity and technique, for example magnetic swipe cards, radio frequency transmitters, transponders and tags relying upon optoelectronic decoding. The fundamental problem regarding tags is that the user is issued with a tag and thus the tag may be used by an intruder who has illicitly acquired the tag; there is no certainty that the person who uses the tag is the rightful owner[1].

This problem can be overcome by using biometrics to authenticate the identity of a user. Biometrics, in this context, is the measurement and subsequent modelling of some part, or parts, of the human body such as the retina[6], fingerprint[8], face[9] and, as will be investigated in this work, human odour. The use of biometrics can greatly improve the reliability of an access control system as each human is believed to be genetically unique[1]. However, this method of access control is reliant upon the accuracy of both the model applied and the sampling and processing of the human feature used [10].

The remainder of this chapter will discuss each concept of an access control system, as previously illustrated in figure 1.2, in more detail.

1.1 User Interface

The user interface allows the user to provide information to the system so the authentication of the user can be assessed. User interfaces can take many different forms but can grouped into the following categories[11]; Tag Based Interfaces, Restricted Information Interfaces and Biometric Interfaces.

1.1.1 Tag Based Interfaces

Tag based interfaces utilise a tag issued to the user which be must be presented to the interface. Unique information is stored within the tag which enables the tag holder to be recognised as a specific user of the system. A large number of tag based interfaces exist ranging from well established mechanical keys [5] and magnetic swipe cards [12] to more sophisticated electronic techniques such as smart cards [13] and radio frequency transponders[14].

Ideally the information stored within the tag should not be able to be altered or copied preventing an intruder gaining access with a 'clone' tag. In practice the data integrity of the tag is dependent upon the technology used; many new tag technologies are soon copied [15] and so provide impostors with the counterfeits to facilitate unauthorised access. Tag based technologies, however, suffer a fundamental flaw regardless of the technology used; information is supplied to the authentication unit from the tag and not the user himself. The tag may be used by anyone who happens upon it, perhaps an intruder. This problem can be overcome by using biometric data to assess whether a user has been authorised.

1.1.2 Restricted Information Interfaces

Restricted information interfaces use a code supplied to, or chosen by, a user; this code is intended to be restricted for use by the specified user only[11]. The most common examples of this method of access control are passwords and PINs (Personal Identification Numbers). Ideally the restricted information should be stored in the user's memory and recalled upon demand. This method of access control is extremely common but used mostly as a method of improving the deficiencies of tag based systems by associating a tag with specific restricted information.

Restricted information based access control systems have the advantages of tag free access and also a secure method of storing information, in a user's memory. However, this method of access control does suffer from the following deficiencies [16]: the information is often written down in order to aid memory, the information is passed onto another user intentionally, the information is observed unintentionally whilst a user is entering it and the user may forget the restricted information. A recent survey[17] revealed that 18% of the people surveyed had been refused cash, from an ATM(Automatic Telling Machine), because they could not remember their PIN. The same survey also revealed that 18% of the people surveyed wrote down their PIN to aid memory, 38% of this group wrote their PIN down directly without applying any personal encryption.

1.1.3 Biometric Interfaces

Biometric interfaces make use of unique human physiological or behavioural characteristics to provide information to the system. Examples of biometric systems range from physiological fingerprints[8], retinas[6, 16] and, in this work, human odour to behavioural characteristics such as handwriting[11] and voice[18].

In many cases two distinctly different types of interface are integrated to form a hybrid system in order to provide a higher degree of security, for example a user may be required to enter a personal identity number (PIN) and provide a fingerprint before access is granted.

This method of access control is reliant upon several factors :

- the accuracy of both the model applied to the human feature
- the reliability of the sampling process of the human feature
- the actual genetic uniqueness of the feature

As a result of the various practical limitations of biometric based access control systems, standard measures of reliability are used in order to compare the performance of different human features and methods of sampling/processing.

The reliability of a secure area can be substantially increased with advent of biometric devices since human characteristics are thought to be unique in nature and are difficult to replicate. Systems involving biometric attributes should ensure that the biometric feature is supplied from a 'reputable source'. For example, iris scan devices must ensure the iris is living by recording reactions to varying intensities of light otherwise a colour photograph could be misinterpreted for a living iris. Biometric systems, therefore, must confirm that the biometric feature is not fake and then must determine, or verify, the identity using the biometric feature.

1.1.4 Ease of use

The transition of authorised personnel into a secure area should ideally not hinder the user in any way. However, in practice the user will experience some delay in using the chosen access control system [2]. For example, instead of merely opening

an uncontrolled door, the user must offer a tag to a reader and wait for the access control system to grant access.

The inconvenience of an access control system can be attributed to two main causes: the usage of the user interface and associated method at the access point, and the delay whilst the system processes the information supplied by the user. The frustration caused by this inconvenience can thus be expressed as a function of the delay:

$$frustration = f[delay] \tag{1.3}$$

where

$f[delay]$ is a function of time spent by the user presenting the information to the system and then waiting whilst the system processes the information, hence:

$$frustration = f_{usage}[t] + f_{processing}[t] \tag{1.4}$$

Frustration will be non-linear since a user would tolerate a limited delay but would soon become frustrated as time progresses. The initial form of frustration could be estimated as an exponential function. However, frustration will saturate at a specific frustration level when the user either retreats from the access point in failure or when the user realises a long wait is necessary for access.

Inconvenience may also be attributed to the psychological fear or discomfort when using the access control device. For example, many people in a recent survey would not even try a biometric device [19] for fear of personal injury as a result of invasive techniques, such as retina scanning. This unwillingness to accept biometric technology may be attributable to various causes, such as fear of the physical effect of the devices and also dislike of the invasive nature of this technology[20].

Concerns have also arisen regarding Human Rights if biometrics becomes widespread[15]. Biometric data initially supplied for access control purposes may be sold for use in different applications. For example, retina data may be sold to an insurance company for medical assessment. Biometric data must remain the property of the individual so that Human Rights can be preserved[15] and biometric data is not abused. Several Christian organisations have also compared the widespread use of biometric technology to various biblical passages regarding the labelling of human beings by use of numbers[21].

These doubts and fears may reduce the progress of biometric technology in the future. Humans are already accustomed to using their features and behaviour for recognition purposes, for example, signatures and human verification of facial features, but these techniques are not automated and do not store human data which could be manipulated by others. However, users may have to adapt as this newer, more secure technology progressively replaces accepted techniques, such as swipe cards and PINs.

1.2 User Authentication

User Authentication is concerned with determining whether a user is an authorised member of the access control system. Authorised members are referred to as subjects where the set of all subjects for a specific system is defined as S. The set of all users is defined as U and contains all people who are subjects of the access control system. The set of all impostors is, therefore, the complement of the user set which can be defined as U', where $U \cup U'$ is the set of all people who are able to attempt access.

The authentication process transforms a user, u, in the real world into a subject, s, of the access control system, which can be defined as :

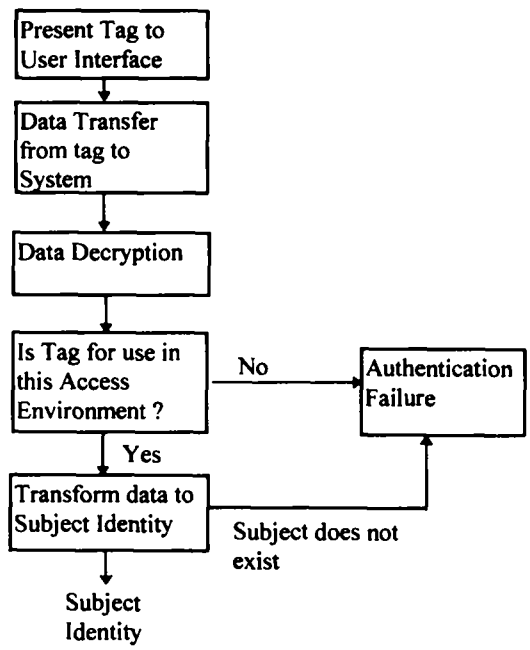
$$user_{auth} : U \rightarrow S \tag{1.5}$$

The interpretation of $user_{auth}(user) = s_i$ is that a *user* is transformed into a subject number *i* by the user authentication function $user_{auth}$. If the $user_{auth}$ function fails to authenticate then an empty set, $\varnothing$, will be returned. The operation uses the information supplied by the user via the user interface. The information supplied may be a tag number in the case of a tag-based system or may be a data set supplied by a biometric interface. Regardless of the type of the input to the module the same output is required: the user is subject, s_i, or the user is not recognised.

1.2.1 Tag and Restricted Information Authentication

The transformation from user to subject may involve many different processes when tag based systems are considered. The data which is stored on the tag must be transferred to the access control system using the user interface, alternatively called tag reader. Once transferred the data must be authenticated for use with the specific access control system and then transformed into the corresponding subject identity, s_i. A diverse range of tags is available which provide differing levels of transmission and data security. For example, a smart card may transmit a pre-encrypted subject identity code using data encryption to avoid an intruder from re-using an intercepted data stream. Conversely, a magnetic stripe card may store and transmit a subject identity directly, transformation to the subject domain does not require any further processing. Figure 1.3 illustrates an authentication procedure which can be applied to tag-based technologies.

Figure 1.3 Tag-Based Authentication



An authentic tag may fail the authentication process due to damaged data on the tag, transmission corruption and errors or decryption / transformation error. The probability of a tag being falsely rejected is dependent upon the specific tag under consideration but should be low.

The probability of a tag producing a false subject identity is high since the tag will produce an identity irrespective of whether the rightful user is in possession of the tag. This probability can be reduced by the introduction of restricted information to verify the rightful user is currently in possession of the tag. For example, a four digit PIN number is required when using an ATM. A survey conducted in 1993[17] revealed that the combination of a tag and a four digit PIN resulted in 1 in 20000 chance that an impostor would be falsely accepted at an ATM. This two stage process of tag and restricted information is slow and increases frustration, especially when large numbers of people require access through such an access point.

Restricted information can be used in isolation without being part of a tag based system. However, passwords or PINs would need to contain a considerable number of digits if large numbers of system users are involved. For example, if an access control system possessed 2000 users and a PIN interface, incorporating a different PIN for each user, then the probability of an impostor correctly guessing a PIN is 0.2 since there are only 10,000 combinations of a four digit numeric PIN. The number of combinations of an N digit password, where each digit has m combinations, is given by m^N. A six digit alpha-numeric password would reduce this probability considerably, to 9×10^{-7}, since the number of possible password combinations would be 2.18×10^{18}. However, these probabilities do not consider the previously discussed disadvantages with restricted information; users forget passwords and PINs especially when they are long and restricted information is easily passed to impostors through written notes or observation.

1.2.2 Biometric Authentication

Biometric systems are more complicated since they supply a probability that the user attempting access is an authorised access control subject. Biometric data is transferred across the user interface so authentication can be performed. This biometric data is compared to subject templates which have been previously loaded into the system. These subject templates are loaded into the system during *enrolment*. Subjects must firstly be created, before templates can be assigned, by means of the following creation function[22]:

$$cr_{subject} : S \rightarrow S \tag{1.6}$$

This create function allows an existing subject of type S to create a new subject of type S for use on the access control system. For example, the interpretation of $create(s_i) = s_m$ is that an existing subject s_i creates a new subject s_m, where i is the

subject number of the creator and m is the subject number of the new subject. The subject performing the creation must have the privilege to create new subjects, this privilege is designated as the administrator privilege.

Once a subject has been created a series of biometric representations, defined as a set of templates T, of the corresponding user can be allocated using the following function :

$$enroll_{user} : U \times S \rightarrow T \tag{1.7}$$

The interpretation of $enroll_{user}(user, s_i) = \{t_0, t_1, \ldots, t_N\}$ is that the $enroll_{user}$ function allocates the N templates, t, supplied by *user* to the ith subject s. One template is sometimes only necessary but usually a number of templates are taken so that a user's natural variability can be determined[11]. This natural variability is especially important when behavioural characteristics are used, for example, a person's signature is assumed to vary each time it is written but this variability must be assessed so that a forgery can be detected.

Biometric systems cannot produce a definite user number corresponding to the current access attempt. Since the uniqueness of the biometric attributes is dependent upon the mathematical representation, the system can only provide a probability, P, that the biometric features supplied by the current user match an enrolled subject. The determination of probability, P, can be represented as :

$$P_{subject} : U \times S \times T \rightarrow P \tag{1.8}$$

Therefore $P_{subject}(user, s_i, \{t_0, t_1, \ldots, t_N\}) = \{p_0, p_1, \ldots, p_N\}$ would return the probabilities obtained when a *user* is compared to the N templates belonging to the ith subject. Therefore, the *user* has a probability of $P(i)$ of being subject i, who is an enrolled subject of the access control system. The authentication system must

apply a threshold, τ, to this resultant probability, $P(i)$, which can be represented mathematically as :

$$f(P(i)) = \begin{cases} True & if\ P(i) \geq \tau_m \\ False & if\ P(i) < \tau_m \end{cases} \quad (1.9)$$

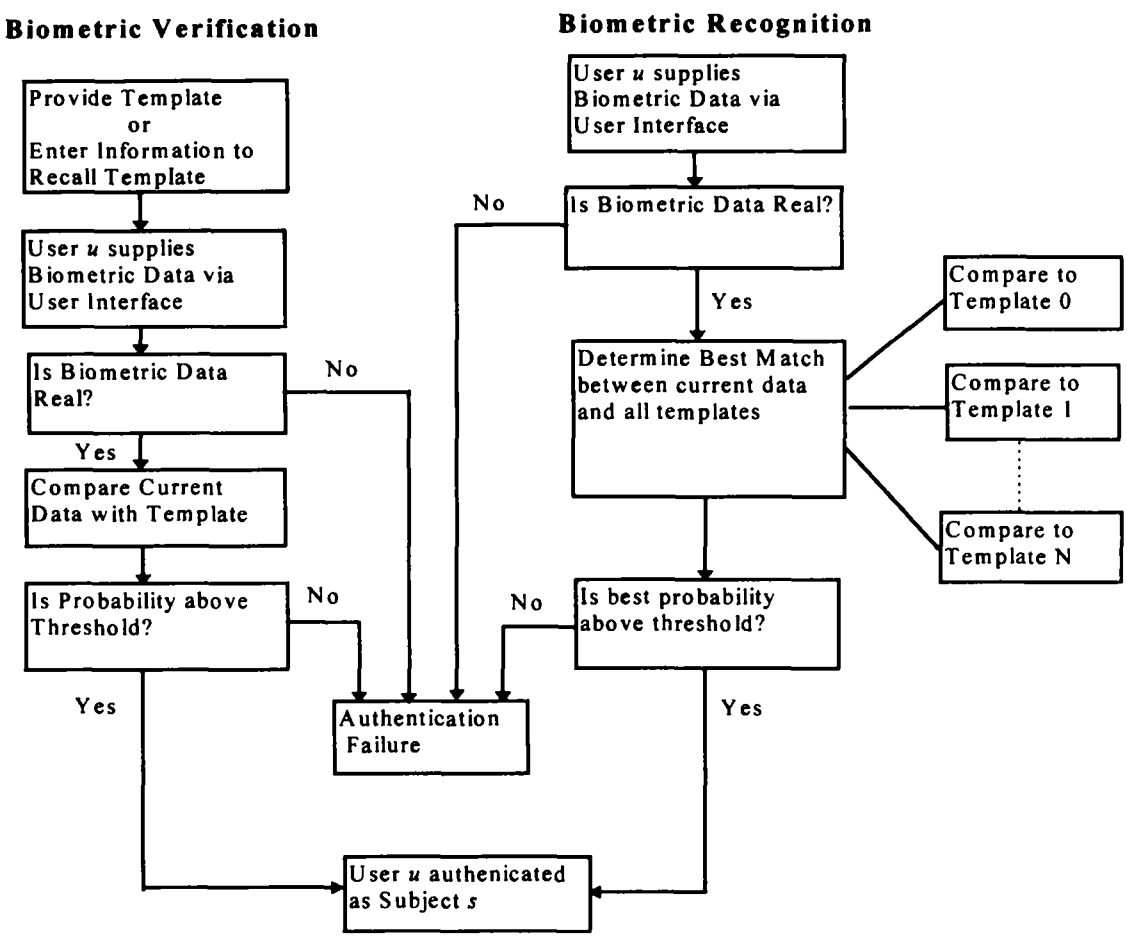
where

$P(i)$ is the probability that the candidate is user i

τ_m is the acceptance threshold at access point m

There are two methods used for processing the biometric feature in use: verification or recognition[23]. Both methods require the user to enrol onto the system by supplying the relevant biometric feature to the system for use as a template for future comparison in real time. Verification methods require the user to enter a personal identification number (PIN) or use a tag in order to call up, or provide, the user's template for comparison with the sample to be supplied in real time. The current sample is then compared with the template and a probability of similarity is expressed, as previously discussed. Recognition systems do not require the use of any tag or PIN and only the biometric feature in question. The real time sample is compared with all templates stored and the chosen subject is the subject with the highest probability. Figure 1.3 illustrates the two methods of biometric authentication.

Figure 1.3 Biometric Authentication



Biometric recognition is a more demanding than verification since the probability of class membership of all system users must be considered as opposed to only one when verification is used. The probability of a false accept in a single verification increases proportionately as the number of users increases[24]. Therefore, the likelihood of an impostor resembling a legitimate system user increases as the number of impostors increases. If $P(v)$ is the probability of a false accept when verification is used then $P(r)$ is the probability of producing a false accept in a recognition attempt involving N impostors can be expressed as [24] :

$$P(r) = 1 - (1-P(v))^N \tag{1.10}$$

For example, if $P(v) = 0.001$ and N = 200 then the probability for a false accept, $P(r)$, increases to 0.181. If the number of impostor attempts increases to N = 2000 then the probability of producing a false accept, $P(r)$, rises to 0.86. Hence, unless the False Acceptance Rate is very low, $P(v) \leq 0.0001$, the probability of allowing entry to an impostor increases to unacceptably high values when using biometric recognition as opposed to verification.

Biometric recognition may also prove impractical as a result of time constraints when determining multiple probabilities. For example, if a biometric system used a one to one assessment algorithm which lasted 0.5 seconds then if the system contained 1000 users then 500 seconds would be needed to perform an exhaustive search for recognition purposes. However, if verification was used then only a single assessment, lasting 0.5 seconds, would be necessary. The time period for performing an assessment is dependent upon the speed of the processing machine, the efficiency of the algorithm, the template size and the type of algorithm.

Several measures are used in order to compare the performance of different biometric devices notably False Rejection Rate (FRR) and False Acceptance Rate (FAR) [19, 23]. FAR can be defined as the probability (expressed as a percentage) that a biometric verification device will fail to reject an impostor. An impostor can be defined as a person who submits a biometric sample in an intentional or inadvertent attempt to replicate another person who has been enrolled legitimately onto the system. FAR is calculated as :

$$FAR = \left(\sum_{i=0}^{n} FA_i \right) . \frac{100}{n} \quad \% \tag{1.11}$$

where

FA is a false acceptance for the *i*th sample.

n is the total number of access attempts.

FRR can be defined as the probability (expressed as a percentage) that a biometric verification device will fail to recognise the identity of a user who is enrolled onto the system. FRR is calculated as :

$$FRR = \left(\sum_{i=0}^{n} FR_i \right) . \frac{100}{n} \quad \% \tag{1.12}$$

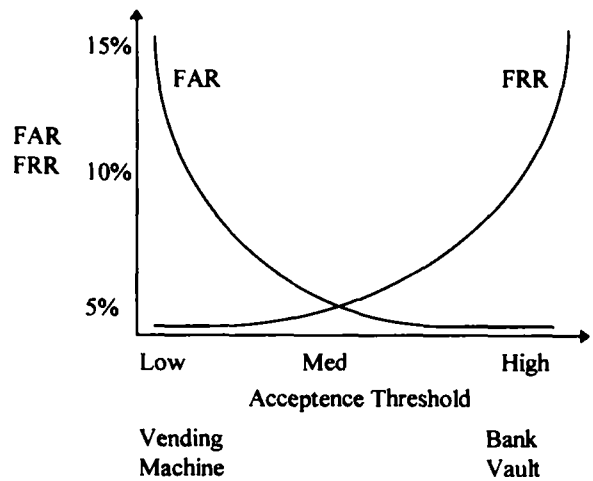
where

FR is a false rejection for the *i*th sample.

n is the total number of access attempts.

These two measurements can be related to the previous discussion of reliability; for a 100% secure system both FAR and FRR must be 0%. Since ideal values of FAR are FRR are not obtainable in a real system, an appropriate threshold must be chosen to provide the most beneficial combination of FAR and FRR. The combination of FAR and FRR is dependent upon the application, for example, a bank vault needs to have zero false accepts because the consequences of an impostor being falsely accepted is high, money could be stolen. However, in order to provide such a low level of false accepts the system may sometimes falsely reject authorised users; the high FRR is therefore as a consequence of setting the threshold to a obtain a low FAR. This concept is illustrated in figure 1.4. Conversely, a vending machine needs to have zero False Rejects since a user will not be willing to supply multiple biometric samples for such a low risk application [1]. A vending machine could tolerate a degree of false accepts since the consequence of unauthorised access into the secure area is low; this would enable the system to operate with low false rejects, as illustrated in figure 1.4.

Figure 1.4 Typical FAR/FRR Curves for a Biometric System



A system which does not exhibit recognition performance may be used as a verification device as an appropriate threshold may be chosen to ensure reliability. However, this increased level of verification performance will also result in a reduction in either FAR or FRR, which may be permissible depending upon the application.

If a biometric access control method is time consuming then a strategy can be applied to the premises which varies the acceptance threshold, and hence the time taken for a decision, depending upon the level of security applied to a specific access point. For example, a high threshold and associated high confidence may be used for an external entrance but then could be reduced once the user has initially gained access to the building under stricter conditions.

1.3 Access Permission Assessment

This section analyses the procedure taken to determine whether a subject should be granted permission to access a secure area. The subject, *s*, who requests permission to access is an authenticated user and is a member of the set of all subjects, *S*, defined for a particular access control system environment *E*. A secure area is accessed via an object, *o*, which is a member of the set of all objects, *O*, defined for the access environment, *E* [4]. An object for a physical access control system is an access point which provides a gateway into the secure area via a barrier.

1.3.1 Access Control Model

The access permission assessment process determines whether the authenticated subject, *s*, has the *permission* to gain access to the secure area via the elected object, *o*. Permission here is not a persistent value but is calculated dynamically by the access control system [4] for each and every access attempt using a number of factors, such as time and day of week, access path chosen to reach door. Permission is a Boolean function since the Access Request Process must either Permit access to the secure area or conversely refuse access. Permission can therefore be expressed as[4] :

$$permission = B(s,c,o,S,O,E) \tag{1.13}$$

where

B is the Boolean function which determines permission.

s is the authenticated subject corresponding to user *u*.

c is the operation which has been requested by the subject.

o is the access object which is the door and consequently the secure area.

S is the set of all subjects.

O is the set of all objects.

E is the environment in which the subjects and objects interact.

An access control system contains subjects who require access to objects. In the case of a physical access control system S represents the set of all people, subjects, and O represents the set of all objects, access points, which the system possesses. Creation of subjects was defined in section 1.2.

Objects must also be created so an object creation function can be defined as[22] :

$$cr_{object} : S \rightarrow O \tag{1.14}$$

Each object possesses a set of rules, R, which are used when determining whether access should be granted to a subject. These rules are created using the following function[22] :

$$cr_{rules} = S \times O \rightarrow R \tag{1.15}$$

The interpretation of $cr_{rules}(s_i, o_m) = \{r\}$ is that the ith subject, s_i, creates a set of rules, r, for the mth object, o_m. The rules may be any factors which influence the access decision process; for example, time limits could defined for an object so that access could only be permitted at certain times. Rules influencing the permission process will be discussed in section 1.3.3.

Once subjects and objects have been created a mechanism must be defined so that object privileges, Pr, can be granted to subjects. A privilege grant function can be defined as[22] :

$$gr_{privileges} : S \times S \times O \rightarrow Pr \tag{1.16}$$

For example, $gr_{privilege}(admin, s_i, o_m) = \{valid\}$ illustrates that a subject possessing the administrator privilege, *admin*, grants the *valid* privilege to subject s_i at some object, o_m. The valid privilege allows entry to the object so long as any other rules allocated to the subject do not result in permission being denied.

If a basic access control system is considered then the previously expressed Boolean permission function may fully describe all of the possible access restraints and may also provide an effective way to manage access through the access points. An Access Control List (ACL) could be used for such a system[4] and would necessitate a separate list for each object, door, which would contain a list of all subjects who will have the *privilege* to access the object. Privilege here is a persistent value which represents whether a user can access an object when all other factors are *not* considered.

Hence, in a basic system the permission to access the object is a direct transformation from the privilege information, which can be expressed as the following rule[25] :

$$permission(s, o) \Rightarrow s \in ACL(o) \neq \varnothing \tag{1.17}$$

where

ACL(o) represents the Access Control List for object *o*.

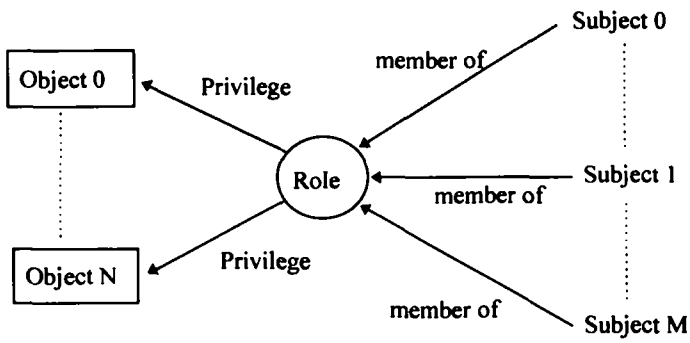
This permission function returns true if the current subject, *s*, is a member of the *ACL* for the requested object, *o*, or false if the subject is not a member of the *ACL*.

1.3.2 Role Based Access Control

When access control systems involve large numbers of subjects and objects a more efficient method is needed to represent the system[25]. Using a Role Based Access Control model greatly simplifies the structure of the system by associating access privileges and restrictions to roles or groups within an organisation[26]. Users with the same role within a company, for example Software Managers, are assumed to require the same access privileges in order to carry out their job function. Therefore, all users within a specific role are allocated to that role within the access control system. Considerable effort would be required initially in order to partition

the roles within the system so that maximum benefit is gained. The relationship between subjects and roles is illustrated in figure 1.5.

Figure 1.5 Relationship Between a Role, Subjects and Objects



Role based systems require careful consideration when determining access privileges for defined roles. Care should be taken to ensure that only the privileges required to perform a specific role are allocated to that role and no more. This criteria is termed the principle of least privilege[25] and ensures a subject does not access a secure area for which he has no need and hence could be deemed an intruder.

A role can be considered as a template subject which is used for permission assessment for all subjects who have been allocated that role. Therefore, roles can be created and configured using similar functions to those used for subjects. Role creation can be defined by :

$$cr_{role} : S \rightarrow Roles \tag{1.18}$$

where

Roles is the set of all Roles with an access control system.

For example, a subject with the administrator privilege creates a new role called 'Software Manager' using $cr_{role}(admin)$ = *'Software Manager'*. The new role can then be configured using the previously defined grant privileges, $gr_{privileges}$, function and create rules, cr_{rules}. For example, $gr_{privileges}(admin, Software\ Manager,$

Front Door) = *{valid}* would allow the 'Software Manager' access to the 'Front Door'.

When a new subject is created a single function can then be used to completely define the new user's privileges and rules. This single function is a grant roles function which can be defined as :

$$gr_{roles} : S \times S \rightarrow Roles \tag{1.19}$$

For example, the interpretation of $gr_{roles}(admin, s_i)$ = {Software Manager} is that a subject, who has been allocated a role which contains the administrator privilege, authorises subject *i* to use the role of 'Software Manager'.

The determination of the Boolean permission function is now modified to account for the incorporation of roles as follows[25] :

$$permission(s, o) \Rightarrow access(pg(rg(s), o))) \tag{1.20}$$

where

rg(s) is a function which provides the role granted to subject *s*.

pg(r,o) is a function which provides the privileges granted to the specified *role* for object *o*.

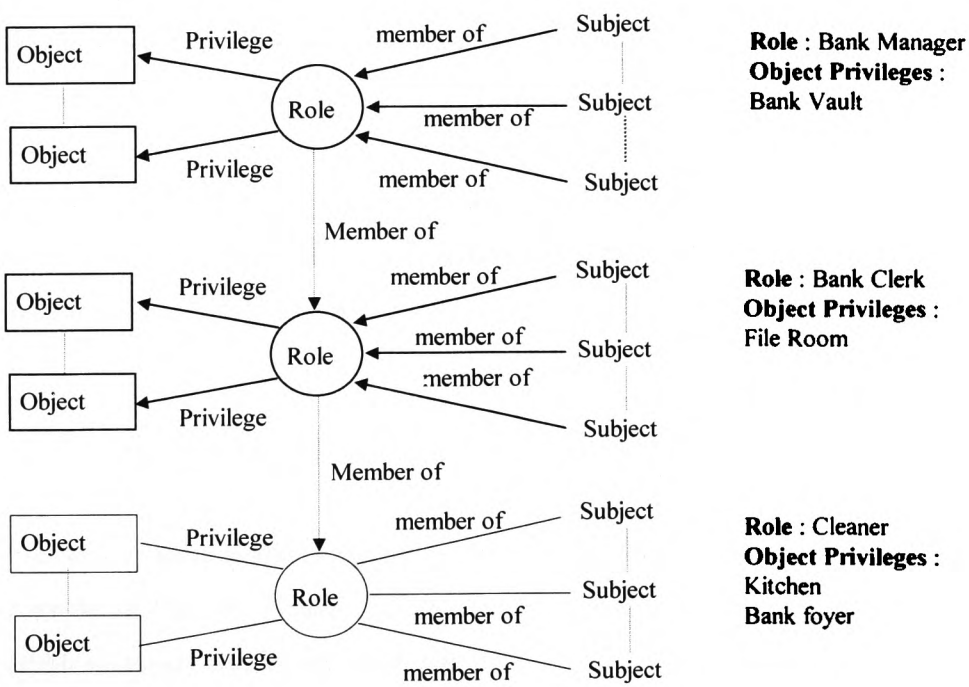
access is a function which determines whether the privileges for the role allow permission to the object.

Once roles have been established it becomes easy for an administrator to configure new system users[25], a new subject is created and a role is granted to the subject. Changes of access privileges also becomes simple since the appropriate new role of a user can simply be allocated and the old role removed.

1.3.2.1 Hierarchical Roles

There are usually certain secure areas within an access control system which are accessible to all users regardless of their allocated role or group. Examples of common areas could include an entrance hall through which all members of staff must pass or a welcome screen on computer software. If common areas exist then a hierarchy of roles can be constructed to avoid duplication of privilege information amongst roles and also to extend the administrative control of the system[25]. As seniority increase then roles could be progressively accumulated. The user is then allocated a number of roles for which he possesses all relevant privileges. This concept can visualised as a hierarchy of roles where roles are defined with increasing security level, figure 1.6 illustrates the role hierarchy architecture and an example when applied to a bank[25].

Figure 1.6 An Example Role Hierarchy



The previously defined function to grant roles to subjects can still be used for hierarchical role methodology but multiple roles can be allocated to a subject instead of just one. Using the grant role function, multiple role allocation can be specified by :

$$gr_{roles}(s_m, s_i) = \{role_0, role_1, \ldots, role_N\} \quad (1.21)$$

where

s_m is the subject granting the roles.

s_i is the subject receiving the roles.

$role_x$ are the roles granted to the subject.

For example, using the grant roles function a Bank Manager would be granted roles as follows: $gr_{roles}(admin, Bank\ Manager) = \{Cleaner, Clerk, Assistant\ Bank\ Manager, Bank\ Manager\}$. Therefore, the Bank Manager holds the privileges for all members of staff which enables access to all areas of the bank.

When role hierarchies are used a modified permission assessment procedure must be performed. The permission is determined for each role granted to a subject s until either the permission for a role evaluates as true or all granted roles evaluate as false, in which case the permission function returns false. The previously defined roles granted function, *rg(s)*, therefore returns all roles currently granted to subject *s* [25]. This can be specified as :

$$rg(s) = \{role_0, role_1, \ldots, role_N\} \quad (1.22)$$

The access permission procedure is more complicated when role hierarchies are considered but the system becomes more powerful and the burden upon the administrator is considerably reduced.

1.3.2.2. Object Aggregation

Objects, for example doors, can be aggregated into groups in similar manner to subjects allocated to roles. An aggregation of objects can be defined as a *zone* in order to clearly differentiate between subject and object aggregation. When a number of objects possess the same restrictions then they may be allocated to a zone and the collection within the zone can be referenced as a single entity. For example, all doors leading to a common secure area could be regarded as one zone which provides access to the same secure area. This condition may not necessarily apply in a practical situation since objects may themselves reside in differing levels of secure area which require differing roles or privileges to enter.

1.3.3 Access Transforms

Transformation of access rights can be considered to be an event which directly influences the verdict of the access permission procedure. This transformation may be external or internal in nature. *External* implies the intervention of a subject, who possesses the appropriate privilege, whereas *internal* transformation implies that the access control system can itself transform the permission verdict, as a result of pre-configured mechanisms[22].

External access right transformations are somewhat limited to the case of object privilege transformations, performed by a subject possessing the administrator privilege. This transformation has been previously defined as the $gr_{privilege}$ function but is modified here to directly incorporate the required administrator privilege :

$$gr_{privileges} : S \times S \times O \times P \rightarrow P \tag{1.23}$$

The interpretation of $gr_{privileges}(s_i, s_n, o_m, \{x\}) = \{y\}$ is that a subject, s_i, possessing the privilege x can grant the privilege y, for object o_m, to subject s_n. For example,

$gr_{privileges}$(*person01, person02, Main Door, {administrator}*) = *{valid}* implies that *person01* can grant the *valid* privilege, for the *Main Door*, to *person02*. This transformation can be termed as amplifying since the receiving subject gains privileges following the transformation[22]. Conversely, the transformation may be attenuating meaning that the receiving subject loses privileges following the transformation. An example of the necessity for such an operation is in case of tag loss or initial set-up of the system.

Internal transformation of access rights occurs when the access control system effects the access permission process as a result of previously configured rules. When physical access control is considered these rules can be classified as time restrictions or route restrictions. These rules may be applied in many different ways depending upon the application, the following possible rules will be considered:

- Anti-pass back
- Route validation
- Time restrictions

1.3.3.1 Anti-Pass Back

Anti-pass back control is a route based internal transformation method. This method only grants *entry* permission to an object if the subject has previously *exited* from the same object. This method obviously requires directional control of the object; for example, a building may have a controlled door which requires permission to enter and exit. The primary use of this method is prevent an unauthorised user gaining access to a secure area because a tag has been 'passed back' through the boundary by some means. This method may also be applied to zones of objects as well as a single object. This internal transformation can be specified as :

$$itrans(s_i, o_m, \{p\}) = \{p`\} \quad (1.24)$$

where

p is the privilege subject s_i currently holds for object o_m .

$p`$ is the new privilege granted as a result of the internal transformation.

For example, the interpretation of *itrans(person01, Front Door, {entry}) = {exit}* is that *person01* obtains the privilege to *exit* the *Front Door* if he enters the *Front Door* first, the *entry* privilege is withdrawn in this case. Access permission to such an object is therefore dependent upon the mode in which the object is accessed; for example, if the object mode is entry then the subject must possess the entry privilege for that object.

1.3.3.2 Route Validation

Route validation involves tracking the route a subject takes to reach an access point. Route validation may be used for various purposes; for example, to ensure that an impostor has not gained access to the secure area via a vulnerable point to avoid detection, or a premises may require passage through several different departments in order to authorise entry. The internal transformation function must be modified to account the involvement of an extra object :

$$itrans : S \times O \times O \times P \times P \rightarrow P \quad (1.25)$$

The interpretation of $itrans(s_i, o_m, o_n, x_m, x_n) = y_n$ is that a subject s_i holding the x_m privilege for object o_m and the x_n privilege for object o_n will acquire the privilege y_n for object o_n. For example, in order to gain permission to access the computer room a subject must first have gained permission to access the Front Door. A subject's *valid* privilege for the Computer Room is transformed from unauthorised, $valid^U$, to authorised, $valid^A$, following access to the Front Door. This internal transformation can be expressed as *itrans(person01, Front Door, Computer Room,* $valid^A, valid^U) = valid^A$.

1.3.3.3 Time Restrictions

Time restrictions only grant access permission for subjects according to time boundary rules associated with objects. Examples for the need of such control are allowing staff access to secure areas on weekdays only and temporary access to premises for contract staff and visitors. These rules may be cyclical which requires boundary rules to be configured for a specific time period, for example a week, after which the boundary rules are applicable continuously on a cyclical basis. The boundary rules must be configured using the previously defined grant rules, gr_{rules}, function. A subject must be granted the corresponding restricted privilege for the time constrained object using the previously defined grant privileges, $gr_{privileges}$, function. The permission status for a subject requiring access to a time constrained access point must then be determined dynamically depending upon the current time.

For example, a time constrained rule, *trule*, is granted an object, *door*, by an administrator, *admin*, using $gr_{rules}(admin, door) = trule$. A subject can then be granted the corresponding restricted privilege, $valid^{trule}$, using the following grant privileges function: $gr_{privileges}(admin, person01, door) = valid^{trule}$. The permission status is then determined dynamically as follows :

$$permission(s, o, t) = access(pg(s, o), ruleg(o), t)) \quad (1.26)$$

where

ruleg(o) is a function which provides the rules granted to object *o*.

pg(s, o) is a function which provides the privileges granted to the specified *subject* for object *o*.

access is a function which determines whether the restricted privileges for the subject allow permission to the object with possessing rules which are governed by time *t*.

1.3.4 Administration Constraints

The administrator for an access control system must possesses the privilege of an administrator. However, it has been assumed that all administrators are authorised to perform all relevant functions related to an access control system. If further administration restraints are not enforced then an administrator may be able to grant privileges or roles which are regarded as more secure than those of the administrator. This type of access control is termed discretionary[25].

The administration privilege can be extended to improve the integrity of the access control system by providing and enforcing new rules for administrators. Mandatory access control[25] provides this integrity by only allowing an administrator to grant privileges or roles to subjects if they are equal or are lower priority to his own.

1.4 Activator

The resultant Boolean decision obtained from the access permission process is used by the activator to produce the relevant action. If the access control subject is successfully granted permission then the activator must release the barrier to enable the user passage through the boundary into the secure area, figure 1.1 showed the various constituents of an access control system. For example, a door may be opened by the activator allowing access to a user.

The activator, however, must also provide an appropriate response if access permission is refused. This action may be to trigger an alarm or to prompt the user to 'try again'. The activator may also provide other facilities such as the co-ordination of event storage in a database.

1.5 Programme of Work

The programme of work used for this research was divided into the following sub-sections :

1. Research into the area of access control. The area of biometric access control was analysed in greater detail considering the project application.

2. Research into the area of electronic olfactory devices and the sensing of odour compounds. Analysis of human odour and the techniques used to analyse multisensor data obtained from electronic odour sensing systems.

3. The design and construction of a prototype biometric odour sensing device within the limits of the current sensor technology available.

4. Development of a suitable test programme, involving a small number of human subjects, in order obtain test data.

5. Analysis of results in order to gain a greater understanding of the data with the purpose of human discrimination as the main objective.

6. Formulation and application of various pattern recognition concepts and techniques to maximise the discrimination capability of the biometric access control method.

7. Analysis and conclusions.

1.6 Industrial Partner

This PhD was completed in conjunction with an industrial partner, Mastiff Electronic Systems Ltd.[28]. The history of the human odour sensing system, named 'Scentinel' by Mastiff, is documented in this section and also any thesis emmissions in order to protect the commercial interests of Mastiff.

1.6.1 'Scentinel' Electronic Nose History

The original concept of using an electronic olfactory device to discriminate between humans was conceived in 1986 by Dr J. Henderson, Managing Director of Mastiff at the time. Dr J. Henderson approached Professor Findlay and Professor Woodward from the University of Leeds with the concept. Following the financial backing of Mastiff's chairman, Viscount Gough, the research project began at the University of Leeds in 1986 and was given the name 'Scentinel'. Preliminary work was directed at both identifying the odour compounds responsible for a human's characteristic odour, under the direction of Dr B. Sommerville, and also the design and fabrication of suitable odour sensors.

Following four years of research, the odour characterisation work was suspended in favour of sensor development. Conducting Polymer sensors were chosen as the most suitable sensors for detecting the organic odour compounds responsible for human odour. Dr T.Gibson was head of a team of two scientists working on the project. The project continued making progress but initial estimates, made by the University of Leeds, for the production of commercial sensors were over ambitious. Unfortunately, this inability to deliver commercially viable sensors heralded the withdrawal of funding by Mastiff.

A company called Bloodhound Sensors Ltd, who are owned by the University of Leeds, are currently continuing research into conducting polymer sensors, as well as other sensor technologies. Mastiff have a licensing agreement with Bloodhound

sensors so that the 'Scentinel' project could begin once again when the sensors have reached the required standard for a commercial product.

1.6.2 Commercial Constraints

The commercial nature of the research imposed restrictions upon the publication of commercially sensitive material in this PhD thesis. The restricted material was information which was regarded as intellectual property of Mastiff and which could benefit competitors to Mastiff. Hence, the following information has not been included, or is not complete, in this thesis :

- Specific polymer compounds used in the prototype odour sensing device.
- Detailed technical drawings of flow chambers and sensing devices.
- Sensor response data, or derived features, sampled during the field trial.

2. Review of Current Work

This chapter will show both the biometric and tag based access control systems currently available at the time of this work, and will highlight the advantages and disadvantages. The other areas applicable to the literature survey are referenced in the relevant chapters of this report.

2.1 Tag based Access Control Survey

2.1.1 Mechanical

The most traditional mechanical method of access control is the lock and key. Modern locks, and associated keys, are very secure, proving very difficult to 'pick' by a potential intruder. The key/lock combinations are not unique but manufacturers provide thousands of different combinations[5]. A key will require an operating period of approximately several seconds, taking into account positioning, turning and extracting the key. The key is still a very effective means of access control but is impractical when greater flexibility and control are required. For example, if each person in a company possessed an identical key for access, then if one user lost his/her key, every user would have to be re-issued with a new key for a new lock.

Various mechanical access control devices exist which rely upon the user to enter a Personal Identification Number (PIN) to gain access. In a similar manner to the mechanical key, if a common PIN is leaked then a new common PIN must be re-issued to all system users. These systems are unreliable as an entry number may be passed by word of mouth to unauthorised users and may easily be forgotten by the user, especially as humans are expected to remember increasing numbers of PINs for various applications [10].

2.1.2 Magnetic Stripe Cards

These cards contain a magnetic stripe which is attached to one side of a card, for example bank cards. The stripe of magnetic tape contains domains of varying magnetism in order to encode data onto the stripe. The magnetic domains are aligned along the tape in one direction. Standard encoding systems are used for many stripe cards which usually store information regarding both data and clocking bits on a single track. A magnetic flux transition between clocks signifies a '1' [12].

The conventional method of reading the information is by utilising an inductive head. Inductive read-heads detect the rate of change of magnetic field amplitude. The output amplitude of the head is, therefore, dependent upon the speed of the recorded medium past the head; for example, a user swiping the card through a reader. An automatic gain control system must be utilised to process this varying signal strength. An alternative method of reading a magnetic stripe card is by using magnetoresistive sensors. These sensors function by passing a current through the magnetoresistive stripe, and the magnetic field is detected by measuring the change in voltage across the device[12]. This type of reader has the advantage of being speed independent which enables a user to swipe a card at any speed.

There are many disadvantages of using this magnetic stripe technology. Data is limited to a maximum of 226 bytes if all three available tracks are used which may not provide enough data area if encryption is used or the card is required to store biometric data. The cards themselves are very susceptible to magnetic fields and may be easily corrupted by using a strong magnet. The cards are also very well established as a technology and hence equipment exists to copy cards very easily. The card readers are also unreliable; for example, inductive head readers are unreliable when 'fast' swipes are made and both readers require contact to the

magnetic stripe which reduces the lifetime of the cards considerably. Magnetic stripe cards are, however, extremely widespread and are very cheap to produce[12].

Watermark magnetic stripe cards attempt to solve the various problems associated with normal stripe cards[12]. This method uses permanent data that cannot be erased or copied; the data contained on the stripe is unique to the card. Unlike normal magnetic stripe cards, watermark cards contain magnetic particles which are at 45 degrees to the tape. The magnetic domain in one length of tape can point in two directions. A specially designed reader can detect where the tape goes from one direction to another. Watermark cards are an improvement over normal stripe cards but still have similar problems associated with swipe speeds and contact wear.

Weigand magnetic stripe cards also strive to overcome the problems associated with magnetic stripe cards. These tags utilise the Weigand effect where a ferromagnetic wire exhibits a spontaneous change in the direction of the magnetic field when a certain field strength is exceeded[27]. Two rows of vertically aligned wires can be used in coded card applications. The reading head contains the sensor coil which is subjected to a permanent magnetic field. The Weigand wires pass through two separate air gaps and their direction of magnetisation changes in the magnetic field. The resulting pulses have different polarisations which can then be used for decoding [27]. Weigand cards cannot be copied or damaged by magnetic fields. The cards also do not require contact from the reading head, hence, mechanical wear is reduced.

2.1.3 Infra Red

Many different types of infra red access control systems exist which differ greatly in operation but can be broadly divided into two areas: infra red transmitter tags and infra red camera technology.

Infra red transmitter tags either constantly emit infra red light or emit pulsed light to conserve energy. An infra red receiver is situated at the point of access which converts the encoded infra red signal into electrical impulses ready for decoding. The infra red token must be correctly orientated towards the receiver since a focused beam is emitted. The distance from the user to the receiver is variable depending upon the application. The system has the advantage of access being assessed before the user reaches the access point which reduces the time taken. Since the tag is constantly transmitting, and hence using energy, the tag must be recharged daily, unless a button is used to activate the transmitter.

Infra Red camera technology can be used to read optically encoded data from a card. An infra red source can be used to illuminate the encoded data on a tag, for example a bar code[28]. The data is then captured using a line scan or a video camera for decoding. This method again provides non contact operation which both speeds up the access process and prevents wear and mechanical failure of the tag. This method does not require any power source for the tag and will not, therefore, require recharging.

2.1.4 Radio Frequency Identification (RF-ID)

This method of access control uses transponder tags which transmit data when a reader sends an interrogating signal[14]. In order to reduce complexity the transponder tag transmits data at the same frequency as the interrogation signal since frequency translation circuitry is complex and requires significant power levels. Signal interference is avoided by time slicing the interrogation and transponder signals; the transponder only responds once the interrogation signal is completed. RF-ID tags are non-contact which increases the tag lifetime and also cannot be easily accessed since the circuitry and antenna are both embedded within the tag.

RF-ID systems are based upon radio frequency or electro-magnetic propagation between the interrogating reader and the transponder, commonly available tags operating in the range 60KHz to 5.8GHz[14]. The coupling between the reader and the transponders can use either magnetic fields or electric fields. Since this method relies upon RF or electro-magnetic propagation they then suffer from a number of disadvantages: the signals are susceptible to electro-magnetic interference (EMC) and in some applications the tags may become inoperable. The signals may also be scanned by an impostor; after being scanned the RF signal may then be re-transmitted by the impostor to gain access. However, most systems incorporate ciphering or code hopping so that the data is transmitted differently each time the tag is interrogated[14].

Magnetic field systems normally use passive transponders, which allow the tags to operate without a power source and hence have an indefinite life span. Frequencies in the range of 125KHz are used for magnetically coupled systems and tags use an antenna comprising numerous turns of wire around a coil former to collect energy from the reader's magnetic field. Due to the magnetic method of coupling, the range of the transponders is restricted to several inches, depending upon the interrogation and transponder antenna.

Electric field systems have a vastly increased range over magnetically coupled systems. These systems use radio frequency electric fields to induce energy and data between the reader and passive tags. The operating range is dependent upon the frequency used for transmission; for example, range is approximately 10m for a 400MHz system and is less than 1m for a 2.5GHz system[14]. If the tag is active, it contains its own internal power source and then the range of the system can be increased considerably[29].

Transponder tags may be read-only or read-write, read-only tags do not require a battery since the data is non-volatile. Read-only tags usually only require a factory

written integrated circuit, which contains driving circuitry and data, and an associated antenna. Read-write tags require more complexity, for additional drive circuits or volatile memory power, to enable this function and hence may require a battery, and will also be more costly than the read-only tags.

2.1.5 Optical Memory Cards

This method of access control contains an optical strip which is bonded to a suitable tag, usually credit card sizes. The strip uses Write Once Read Many (WORM) architecture which means that data can be written to the tag but can never be erased. Memory capacities of 4.1 Mbytes have been reported[30] which would also enable this type of tag to be used to store biometric data if necessary. The data contained on the tag can be stored using any type of encryption standards, for example DES and private-public key methods, to improve the integrity of the tag. The tag is non-contact since optical methods are used for interrogation but external damage to the tag may result in data loss and consequently refused access. However, this method is completely immune to RF, magnetic or electrostatic interference.

Data is written onto the optical strip using a low-power semiconductor laser which split into three beams: two used for locating and locking onto a track and one used for data writing[30]. The laser energy marks the optical media forming pits, data bits, of approximately 2.25 microns in diameter. The pits can not be replaced, altered or removed once the writing process is complete. The read process still splits the beam into three but the reflected energy is measured to determine whether the data bit is a dark pit or a reflective area.

2.1.6 Smartcards

Smartcards merge the technologies of memory and microprocessors onto a credit card sized package[13]. An on-card microprocessor controls access to the on-card memory which can store up to 16Kbytes of data[31]. Smartcards can be read-only or read-write; the read-only cards must be factory programmed but benefit from reduced complexity. The interface to a smartcard is either via physical contacts which are engaged when a card is inserted into a reader or via remote magnetic or electric field coupling, called contactless smartcards. Contactless smartcards have a longer life time and are also quicker to use where mass transit of people may be needed[31].

Smart Cards do not contain a power source and so rely upon the reader to supply the necessary power for running on-chip software[13]. The card contains non-volatile memory to enable software and data to remain on the card when the reader's power source is disconnected. The user data is stored in EEPROM (Electrically Erasable Programmable ROM), occupying between 1 and 16 Kbytes in size, which enables the data to remain persistent when power is removed but also allows the data areas to be re-programmed[13]. The software which controls access and data transfer is stored in persistent ROM, occupying between 3 and 20 Kbytes, which can not be altered. The final memory is volatile RAM, occupying several hundred bytes, which is used by the smartcard software when transferring data or performing access checks.

Smartcards can be operated in different access modes depending upon the degree of security required for an application. A smartcard initially requires the correct 64 bit pass code[32] which has been pre-loaded into the card; this code may represent all cards allocated to a specific customer site. A 64 bit pass code provides 2^{64} or 1.8×10^{19} combinations which makes the probability of a prospective impostor reading the card extremely low; smartcards only allow one attempt at the pass

code per second to prevent an impostor cycling through all possible codes electronically[32]. A reader may then gain access to user data or may prompt for verification in the form of a PIN or password. Smartcards can also cipher data transmissions to avoid an impostor scanning a code, for the case of contactless cards, and re-using the pass code and user data in a counterfeit card [13].

2.2 Biometric Access Control Systems

2.2.1 Physiological

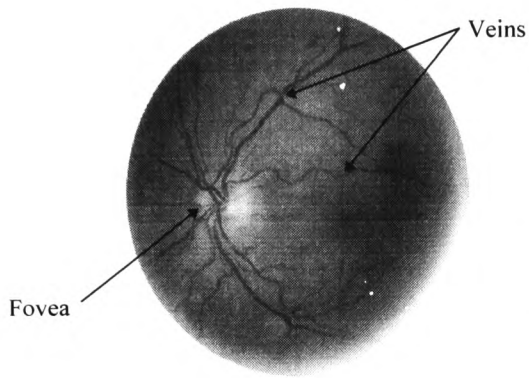
Access control systems under this category rely upon the unique characteristics of a physical part of the human body.

2.2.1.1 Retinal Vascular Pattern

This biometric technique relies upon the fact that no two people have the same pattern of blood vessels in the retina of the eye, as illustrated in figure 2.1; even identical twins have different retinal patterns [16,6]. The retinal vascular pattern is a very stable biometric feature since the eye possesses a stable environment similar to the brain [6]. The vascular pattern will remain unchanged indefinitely unless disease or serious physical injury to the eye is sustained, which could alter the vascular pattern resulting in system failure.

The retinal scan technique works by constructing a map of the vascular pattern on the retinal portion of the eyeball. This scan is performed using infra red light and an associated video camera. The light is directed around the rear of the eye, centred on the fovea as illustrated in figure 2.1 [6], to provide a 360 degree circular scan. This scan is represented by a data set of approximately 200 data points, which supplies enough data to distinguish between any two retinas [16]. The infra red light can be provided by a standard flashlight bulb with appropriate filters, however, some earlier techniques used laser light in order to provide the required accuracy.

Figure 2.1 The Vascular Pattern of a Human Retina



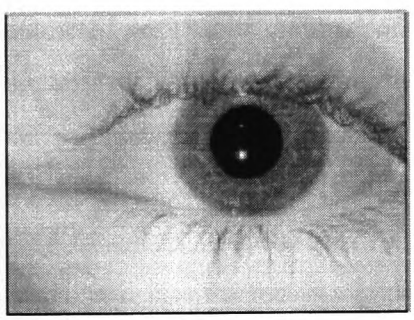
This technique has the advantages of being very reliable; it has a claimed FAR of 0.0001% and a FRR of 0.1% when operated in verification mode[6], dependent upon threshold settings. However, this method does possess some distinct disadvantages wholly associated with ergonomics and the degree of risk, actual or perceived, to the human eye. The system is uncomfortable to use, requiring the user to stare directly into a 'black hole' during scanning. The user is also not aware of any scan taking place since infra red light is used, which is undetectable to the human eye. The use of this wavelength of light suppresses the eye's natural reaction to high intensity light. A user's eye could, therefore, be permanently damaged if the infra red light intensity was not accurately controlled. Users have also objected to the system for hygiene reasons because of the close proximity required between the device and a user's face and eyes [16].

2.2.1.2 Iris Pattern

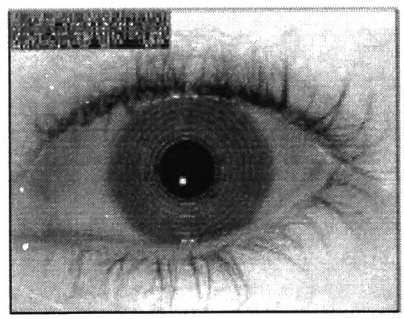
This method is similar in operation to the retina scan technique but instead of using the retina, the iris is used to construct a unique map. Each iris is unique to every eye, with each individual's own two eyes exhibiting differing iris patterns [33]. Identical twins, monozygotic, also exhibit differing iris patterns[33].

The iris is the contractile curtain perforated by the pupil and makes up the coloured part of the eye. Unlike the retina scan technique, standard video optics are all that are needed to capture an image of the iris; the use of infra red or any form of high intensity light is not required. The patterns, that are observable in the iris, are made up of furrows (grooves, wrinkles), crypts (small cavities), corona (coloured rings), filaments (fibres), pits, freckles (discolouration), striations (fine streaks) and rings[33]. A typical iris is illustrated in figure 2.2 a [34].

Figure 2.2a A Typical Iris



2.2b Isolation of an Iris for Encoding and the Resulting Iris Code



Unlike retina scans the user is not required to contact any eyepiece but merely looks into a camera from a short distance, approximately 20 cm [33, 34]. The camera then produces a black and white image which is stored in a 256 byte encoded template resulting in FAR of 0.00076% [10, 34].

A user is able to be enrolled onto the system by means of a photograph. However, whilst on line, the system looks for random movements in the size of the pupil to ensure the subject is alive and not just a photograph. The iris scan technique is more comfortable to use than the retina scan technique but still uses the eye which the user could perceive to be a vulnerable area of the body.

2.2.1.3 Facial Visual Characteristics

Facial recognition relies upon the unique characteristics of a human's face; this is the primary method of recognition between humans. Automation of facial recognition involves several different areas: image capture ensuring equal exposure and frontal orientation, facial detection and framing, facial feature extraction and finally pattern recognition[35]. Once the facial image is captured, compensations can be made to account for orientation and facial inconsistencies such as differing facial expressions [36]; this is called elastic pattern matching since the facial image is stretched back into a full frontal orientation.

Once the facial image is extracted from a scene, the major features must be extracted to enable comparison with previously stored templates. A common method to extract features is to use the Karhunen-Loeve expansion, otherwise known as principal component analysis, which produces Eigen vectors which represent the facial features in descending detail[9,35]. Consequently this method is often called 'Eigen Faces'. Figure 2.3[9] illustrates a number of Eigen faces which are constructed using the first Eigen vector only which describes the majority of the variance of the facial image. Figure 2.3 demonstrates that faces can be visually differentiated even when just one Eigen Vector is used. Approximately twenty Eigen vectors are used for further pattern recognition which may use neural networks [9,35] or statistical techniques.

Figure 2.3 Representation of Faces using First Eigen Vector

A practical face recognition system must allow for various changes in facial features such as orientation, hair style, facial hair. These problems are overcome using data redundancy and also use of texture and wavelength data from the captured image[37, 36]. The authenticity of a face must also be assessed to ensure that the facial image is not a photograph or a latex mask. Several time- separated images can be taken to ensure that the face shows human movement characteristics. Infra red absorption characteristics may also be measured to determine whether the face has been captured for a real living person or just a photograph or latex mask[36].

The system has the advantage of being easy to use as no part of the body is required to contact the apparatus. The system is also able to store 'images' of all attempts on the system; this is obviously advantageous when trying to trace an impostor. The facial verification system has several drawbacks, however, as follows: the data capture sequence can be severely affected by bright light or shadows which make it difficult to perceive edges of facial features or may introduce false edges. Current systems produce inferior recognition rates, approximately 99%, when compared to other biometric technologies.

2.2.1.4 Hand or Finger Shape Profile

Shape-based verification systems rely upon the assumption that each human's hand or finger has unique characteristic dimensions. The relative lengths of fingers and shape/height of knuckles are just two such examples of the features that are used for identification.

The hand or finger to be assessed is aligned onto a measuring grid by means of locating posts situated in between the fingers. A 3-dimensional profile of the hand or finger is then compiled using various images of the hand captured at different orientations [33].

The template taken at enrolment must be able to account for variations in hand position and also variations in hand size caused by heat. The system must incorporate a method of determining whether the hand is real or just a latex/plaster copy of an authorised user's hand or finger. It is believed that this form of authentication may not be truly unique [15] and hence may not be suitable for use in high security access control systems.

2.2.1.5 Finger or Palm Prints

The use of fingerprints as a biometric technique to ascertain an individual's identity is well established and widespread. Fingerprints have been used by police forensic experts for nearly 100 years [8]. An image of the finger or palm print is first captured by a video camera, as illustrated in figure 2.4a[38]. Edge detection is performed on the image which produces a single pixel width image for each ridge of the fingerprint, as illustrated in figure 2.4b[38].

This edge map is then used to extract the key features which represent the finger or palm print. One such method of feature extraction is coincident sequencing,

used by police forces as a manual method for many years. This method ignores the principal finger or palm print features such as whorls, hoops and arches and instead uses the 'minutiae' features. Minutiae are the points on finger or palm prints where ridges end and where they bifurcate (divide). This technique uses relationships between a number of these minutiae features in order to classify a finger or palm print. Figure 2.4 [38] illustrates a processed finger print with several minutae features identified.

Figure 2.4 Coincident Sequencing on a Finger Print

a. Digitised Fingerprint

b. Edge Map of Finger Print

c. Processed Fingerprint showing Minutae

For a coincident sequence to exist there must be at least four features found in the fingerprint under inspection. For criminal identification, in the UK, a total of sixteen features are needed in the coincident sequence. However, for civilian use it is generally accepted that eight features are sufficient in order to make an identification on the basis of finger or palm prints. Even with a full analysis of

seventeen features, required in France for criminal use, using the aforementioned method requires 918 bits of information to be stored per user.

During verification, the processed 'real time' image is used to extract all of the reliable minutiae features, which may be more than the seventeen stored in the template. The relevant data is then calculated for each feature (type, location, orientation and ridge separation). This data is then compared to the data stored in the template until eight similar features are found. If eight similar features are not found then verification is refused. Obviously this system has a large degree of redundancy which will protect against changes in fingerprints caused by cuts, abrasion and dirt.

The disadvantages of fingerprint systems for access control are both physical and psychological. Users do not like the connection between fingerprints and criminal records [15]. Some fingerprint systems may be susceptible to fake fingerprints, for example latex or pigskin. However, a technique called total internal reflection spectroscopy can distinguish a real finger from a fake [33].

Fingerprint verification systems are well established and very good FAR and FRR rates are obtainable; FAR varies from 0.00001-1% and FRR varies from 0.05-4% depending upon the system in use [33].

2.2.1.6 Hand Vascular Pattern

This biometric system uses the pattern of veins in the back of the hand to verify an individual's identity[39, 1]. This method assumes that the pattern of the subcutaneous vein structure on the back of the human hand is unique. An image of the vein pattern is captured using the near infra red spectrum, as illustrated in figure 2.5 [39]. Infra red wavelength light is used to provide a high contrast vein pattern image regardless of physiological variations such as age, sex and levels of

subcutaneous fat. This wavelength is also relatively immune to skin features such as moles, warts, scars, pigmentation and hair[1].

Figure 2.5 Example of Back-of-Hand Vein Pattern

a. Digitised Image b. Edge Map c. Difference between two vein profiles of same person

A vein structure comprises a set of nodes connected by links. The structure can be characterised according to topological and geometrical information and diffuseness. The system can differentiate between a living hand and a forgery by comparing two different patterns: one pattern taken with light at the absorption wavelength of haemoglobin and the other for oxyhaemoglobin [1]. The system can then verify that the subject is alive.

The disadvantages of this system are that vein patterns are very difficult to reproduce as veins are free to move around in soft tissue and there will also be some variability depending on physiological state such as temperature [1]. In addition, the position of veins within the hand is dependent upon how the hand is flexed; in fact, a vein could appear to disappear if the hand is flexed in a particular manner.

2.2.1.7 Facial Thermal Characteristics

This biometric method relies upon the principle that underlying vascular system in a human face is unique to an individual [38,40]. Heat from this network of veins passes through the facial tissue and is emitted from the face as infra red radiation, producing a unique thermal pattern[38 40]. An infra red camera is used to capture the thermal pattern for analysis and feature extraction. The vascular heat patterns generated by the nose and ears are detected and discarded since they are highly sensitive to temperature fluctuations [38]. Figure 2.6 shows how facial vascular patterns differ between monozygotic, identical, twins [38].

Figure 2.6 Thermal Facial Vascular Pattern Differences for Monozygotic Twins

Facial Images

Corresponding Thermal Vascular Patterns

This biometric method has similar benefits to facial feature recognition; it is non-invasive since no contact is necessary and use of the face for recognition is more accepted by users. This system has several further advantages notably its independence to light since infra red emissions are used and the fact that the technology is resilient to impostors since emulation of a subcutaneous vein network is difficult[38].

The system requires 2 to 4 Kbytes of storage for each thermal face map. This method of biometrics is relatively new and so will require extensive testing and field trials to substantiate any claims of high recognition and reliability [38].

2.2.2 Behavioural Biometrics

Access control systems under this category rely upon the unique characteristics of human actions and behaviour.

2.2.2.1 Voice

It is believed that each human's voice is unique and that no two voices will be exactly the same. The voice is unique as it is made up from a variety of unique characteristics which include the size and shape of mouth, vocal tract and nasal cavities [10]. Voice-based systems attempt to extract these vocal characteristics of the user, rather than merely determining what the user speaks.

Voice verification and recognition devices have been the centre of much research over previous years. There are numerous devices available which differ greatly in their signal processing techniques. Since the voice is a behavioural biometric feature, it is prone to human variations. These variations range from inconsistent speech to modified speech due to infection. Noise also presents a great problem to voice recognition systems. This noise may be other human speech or noise from a street corner. A voice recognition system must either eliminate the source of the noise, by situating the microphone in a quiet area, or by extracting the characteristic voice pattern from the noise. Recent trials have proved that voice-based systems cannot cope with noisy environments [37].

Converting the acoustic speech into an electrically varying signal is performed very easily and cheaply by using a microphone. The resulting signal is then

processed to extract the characteristic features of the human voice. One such method is by using the zero crossing points of the speech pattern. The information contained at the crossing points is said to represent a very high proportion of the characteristics of the voice [18].

Enrolment onto voice systems usually involves the user speaking a variety of different words. The system only requires the user to say two to three requested words whilst the system is on line. The system must be able to contend with an impostor who can submit data in the form of prerecorded speech or by impersonating an authorised user. FAR and FRR rates differ greatly depending upon the system in question but range from 0.2-5% for FAR and 0.03-5% for FRR [37, 18].

2.2.2.2 Dynamic Signature Characteristics

This biometric uses numerous characteristics which can be observed when a person's signature is written [19,11,41]. These characteristics can be divided into static and dynamic features. Static features include number of component strokes, ratio of long to short strokes, curvature measurements, segment lengths and space usage[11,41]. Dynamic characteristics include timing measurements, stroke order, pen velocity profile, pen acceleration profile, pen up/down pattern, pressure shapes, stylus friction, tilting and slanting angles [11,41]. The static and dynamic measurements are taken using specialised pads and associated pens which perform the necessary dynamic measurements [19]; for example, pen tip pressure is measured by pressure transducers embedded in the signature pad.

The number of data points needed varies depending upon the number of attributes which are stored but a typical template size is 2 Kbytes[19]. Template size may also change if a dynamic feature extraction is utilised which stores attributes in

proportion to the complexity of the signature[11]. During enrolment, at least three signatures are required; these three template signatures are then compared with each other to determine a personal threshold for an individual. This threshold is needed as it is appreciated that a user is not able to replicate a signature perfectly, therefore, the threshold allows for the natural variability in a human signature[19]. Reports suggest that variations in a person's signature caused by fatigue, illness or temperament can be tolerated [41].

This method suffers from several disadvantages: the method assumes that every user is able to write and is able to perform a characteristic signature. This method involves the user performing a skill which requires a higher level of awareness than most other techniques, such as hand profile and finger prints, which are more passive in nature. Signing a signature to gain access through a door, for example, may become tedious for an access control user and lead to high levels of frustration.

2.2.2.3 Keystroke Dynamics

This is a relatively new concept in biometrics which uses an individual's characteristic typing rhythm to ascertain his or her identity. The system measures the way in which a person types. Each person has certain characteristic features which include dwell time: the time a finger holds a key pressed, and flight time: the time in between successive key presses[37].

It has been claimed [37] that even occasional keyboard users have individual timings between certain groups of keyboard characters. As few as a dozen key strokes are enough to determine a user's rhythm. This method has the added advantage of being able to monitor a keyboard operator's identity continuously with any change in user being quickly detected. However, a FRR of less than 2%

has been documented[42] which is currently too low for use in a physical access control system.

2.3 Discussion

Tag based methods are well established for access control applications and generally offer high levels of reliability. Traditionally some form of user interaction was needed when using tags, for example swiping a magnetic card, which decreased ease of use and consequently frustration levels may have been high. However, recent tag based technologies reduce this deficiency by operating remotely to various degrees. Such systems which incorporate a long range may operate autonomously, effectively reducing frustration levels considerably. This reduced frustration level assumes that the user remembers to carry the tag, and is consequently dependent upon human error. However, tag based systems still have no method of verifying that the user is authorised to use the tag; the tag is granted access irrespective of the user. The use of restricted information improves this shortfall but is prone to human error and increases frustration levels due to increased time periods required for processing a single access request.

Biometric methods potentially solve the main deficiency of tag based systems because a human attribute, physical or behavioural, supplies the authentication data to the access control system. However, despite the prospect of more secure access control, biometric systems have so far failed to penetrate into the access control market to any great extent. This failure can be attributed to system reliability deficiencies and the reluctance of users to use biometric technology. For example, the techniques of retina scanning and iris scanning have been shown to exhibit very high reliability but are very intrusive and hence unpopular with users. Conversely, facial and voice recognition are relatively unobtrusive but have not demonstrated the reliability necessary for a commercial system.

An ideal biometric system would operate remotely and unobtrusively in order to increase ease of use, in a similar manner to the migration of tag based systems to 'hands free' use. The use of human odour as a biometric method may eventually fulfill these requirements since odour can be detected regardless of orientation of the user and does not require any user interaction.

3. Design of an Electronic Olfactory Device

This chapter considers the various disciplines which must be integrated in order to construct an electronic olfactory device, commonly called an electronic nose. Odour is discussed so that an insight can be gained into how different odours combine and diffuse through space. The odour of a human is then addressed with particular focus placed upon the evidence for the uniqueness of an individual's odour and the effectiveness of different odour sources of the human body. The biological nose will be discussed as an insight into the various parts of an electronic nose. Finally the various constituent disciplines will be integrated to produce the design of the electronic nose used in this work.

3.1 The Nature of Odour

Humans use three main senses in the determination of flavour [43], those being gustation (taste), olfaction (smell) and the trigeminal sense (skin and mucous membranes). Gustation processes mainly non-volatile chemicals, usually liquids, whereas olfaction processes volatile gases. The trigeminal senses process both non-volatile and volatile chemicals. Humans often underestimate the importance of smell in their day to day lives; most humans if questioned would chose to lose their sense of smell over any other [43]. However, olfaction plays the most important role in the identification of flavours; for example, the tastes of beer and lager are almost indistinguishable if a subject's nose is held closed [44].

Odour compounds are typically small hydrophobic molecules with relative molecular masses in the range 30 to 3000 Daltons[1]. The manner in which molecular structure and odour type relate to each other is understood in general terms but detailed odour-structure relationships are unknown[45].

[1] A Dalton is a measure of atomic weight where 1 Dalton = 1/12 of the atomic mass of Carbon-12. Daltons are also known as amus (Atomic Mass Units).

It is believed that shape, size and polar properties of the molecule determine its odour properties but exact rules are poorly understood. The operation of a mammalian nose has been linked with these physical properties of a molecule and not actually by the chemical composition of the molecule[46]. This theory explains how two entirely different volatile chemical compounds seem to smell the same.

The threshold for different molecules can vary significantly depending upon the odourant in question. The threshold of an odour can be described as the actual quantity of odour molecules (ppm or ppb) that are needed before the odourant can be detected by the nose, whether mammalian or artificial. Some molecules have very low thresholds which could cause problems in an odour sensing system. A high threshold characteristic odour could be masked by a low threshold contaminant of a similar molecular structure even if the low threshold odour is far more abundant than the contaminant. Many off-flavours are caused by very small amounts, below parts per billion, of these powerful odorants[46].

The gas mixture which is sensed by a mammalian nose or an electronic nose may contain unwanted gases in addition to the analyte gas or gases. These unwanted gases may not affect the resilient odour concentrations detected by a 'nose'. Conversely the unwanted gases may cause interference with the analyte gases producing an inaccurate representation of the analyte mixture. For example, ethanol and formaldehyde interference in the measurement of carbon monoxide[47]. Gases can interfere with the measured analyte in the following different ways :

- Positive interference : where the interfering gas adds to the signal which gives a higher analyte measurement.
- Negative interference : where the sensor response to the interference gas generates a signal that opposes the analyte signal, so producing a lower analyte measurement.

- Reactive interference : where the interfering gas reacts chemically with the analyte in the gas phase, either reducing or increasing the analyte response.
- Activation : where the sensor is made more sensitive to the analyte, even after the interfering gas has gone, so producing a higher analyte measurement.
- Poisoning : where the response of the sensor to the analyte is permanently or temporarily reduced by exposure to the interfering gas.

3.2 Odour Model

Some knowledge of the way in which odours diffuse and interact in the vicinity of the sensor array is needed when considering the electronic nose design. Concentrations of odour substances have been shown to change in a non-linear manner with distance and time[48]. The concentration of odour decreases as the distance increases from the source due to the diffusion of the odour molecules into the atmosphere. The odour concentration also varies as time progresses since the odour concentration emitted by the source can vary with respect to time. Different odours have consequently different diffusion characteristics, therefore relative concentrations of odours also change with time and distance.

A gas relies upon various mechanisms in order to travel from the odour source to the receptor, the sensor array in this case. The mechanism of diffusion is considered here whilst other effects, such as convection, are assumed to be negligible for demonstration purposes. An odour source, of finite size, diffuses spatially and temporally in an infinite space according to the following diffusion law [48] :

$$\partial n / \partial t = D\nabla_n^2 \tag{3.1}$$

where

n is the concentration of the odour molecules as a function of time, t.

D is the diffusion coefficient of the released odour.

If the odour source is assumed to contain a finite number of N molecules the manner in which the odour molecules diffuse is governed by the solution to the diffusion equation, as shown previously. If spherical symmetry is assumed for demonstration purposes the diffusion equation reduces to [48] :

$$\frac{\partial n}{\partial t} = \frac{\partial^2 n}{\partial r^2} + \frac{2\partial n}{r\partial r} \tag{3.2}$$

where

r represents the distance from the odour source after time, t, from the release of the odour at time $t = 0$.

The solution to this equation is given by [48] :

$$n = N(\pi Dt)^{-1.5} \exp(-r^2 / 4Dt) \tag{3.3}$$

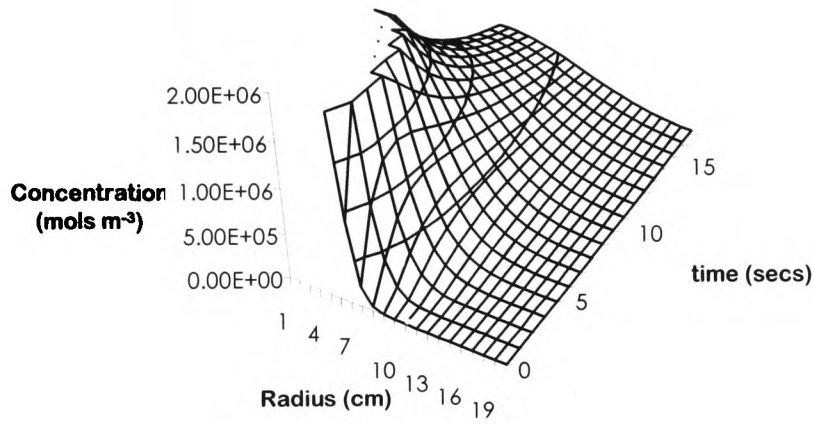
where

N is the quantity of odour molecules contained in the odour source.

The diffusion characteristics of a specific odour can now be calculated if the diffusion coefficient, D, of the odour is known as well as the number of molecules, N, contained in the source. If N is given a nominal value of 100 and ethyl alcohol is chosen as the source, $D = 1.02 \times 10^{-5}$ m^2s^{-1}, then the corresponding concentration map is shown in figure 3.1.

The concentration of odour molecules rises sharply with time from zero to a maximum value which is off-scale in figure 3.1. The concentration can then be seen to reduce rapidly as both time and distance increases. Figure 3.1 illustrates that, if a non-replenished source is considered, the sampling distance from the source and the time the sample is taken both influence the magnitude of the resultant odour concentration.

Figure 3.1 Concentration of Ethyl Alcohol for a Non-replenished Source in an Unbounded Atmosphere.



This result must now be modified to account for the human odour source used in this odour sensing device. Instead of containing a finite number of odour molecules a human can be viewed as a replenished source in which molecules are replaced as they diffuse. A human sweat duct can be considered to release secretion to replace molecules which have diffused into the atmosphere. In order to model a replenished source a small sphere, of radius r', is considered. For a one dimensional case the aforementioned diffusion equation can be reduced to [48] :

$$\partial n / \partial t = D\partial^2 n / \partial x^2 \tag{3.4}$$

The solution can be shown to be [48]:

$$n = (N / r)erfc[(r - r') / 2\sqrt{(Dt)}] \tag{3.5}$$

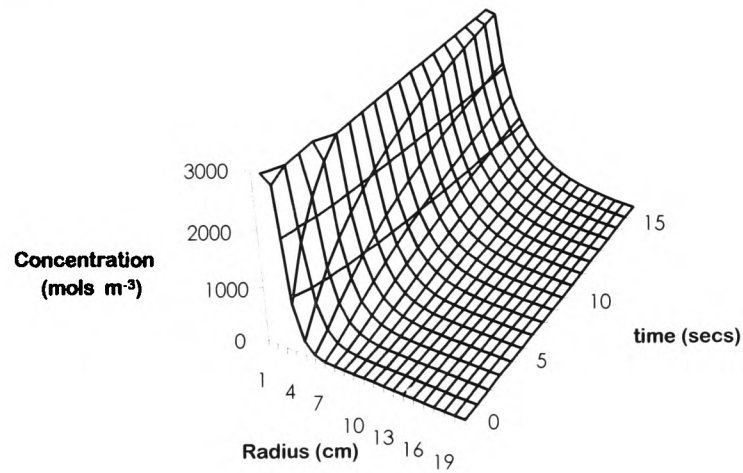
where

efrc is the complementary error function defined as :

$$efrc(x) = 2 / \sqrt{\pi} \int_{x}^{\infty} e^{-t^2} dt \tag{3.6}$$

Figure 3.2 shows the diffusion map for a replenished source, assuming the same odour and constraints as the map illustrated in figure 3.1. The concentration increases sharply from time zero for distances close to the source but more gradually with time as the distance from the source increases. The odour concentration can be seen to fall sharply as distance increases from the source. The replenished diffusion map, shown in figure 3.2, is notably different to the non-replenished map, illustrated in figure 3.1 in that the non-replenished concentration level reduces to a greater extent as time increases. This is caused by the scarcity of molecules as diffusion occurs spatially and temporally. Both maps show that if sufficient time elapses, so that a steady state is attained, then odour concentration can be predicted as a function of distance from the source only.

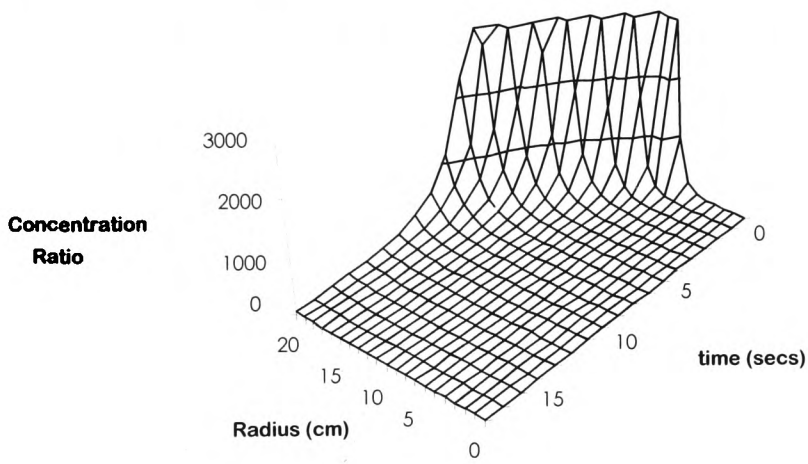
Figure 3.2 Diffusion Map Emitted from a Replenishable Source



These concentration predictions have involved only a single type of molecule which is an extreme simplification of a human odour source. A new odour molecule can be modelled by substituting the diffusion coefficient, D, for the new molecule. Diffusion maps for various substances show no obvious differences

when considered in isolation but relative concentrations of different substances demonstrate considerable imbalances in concentration with respect to time and distance. Figure 3.3 illustrates an example of this concentration ratio for two substances with differing diffusion coefficients, methyl alcohol to ethyl alcohol. Figure 3.3 implies that a source comprising numerous substances will produce odours which continuously vary in absolute strength but also in relative composition. Therefore, an odour from a multiple source may possess a different 'smell' depending upon when and where it is sensed.

Figure 3.3 Concentration Ratio of Two Replenished Gas Sources



Relative concentration imbalances reduce at steady state but are still significant, for replenished or unreplenished sources. Figure 3.4 illustrates the concentration ratio between ethyl alcohol, $D = 1.02 \times 10^{-5}$ m^2s^{-1}, and three other gases with differing diffusion coefficients as indicated; all other parameters are equal for the four gases. Figure 3.4 demonstrates that as the distance from the odour source increases, the concentrations of different gases vary at different rates. Figures 3.4 illustrates that even when a compound of similar characteristics is considered, $D = 1.1 \times 10^{-5}$ m^2s^{-1}, this lighter gas has a concentration 20% greater at 10cm from the source. The increase in concentration is 95% when a significantly lighter gas is considered, $D = 1.0 \times 10^{-4}$ m^2s^{-1} as illustrated in figure 3.4.

Figure 3.4 Concentration Ratio Variation Between Ethyl Alcohol (D = 1.02 × $10^{-5}m^2s^{-1}$) and Three Other Gases

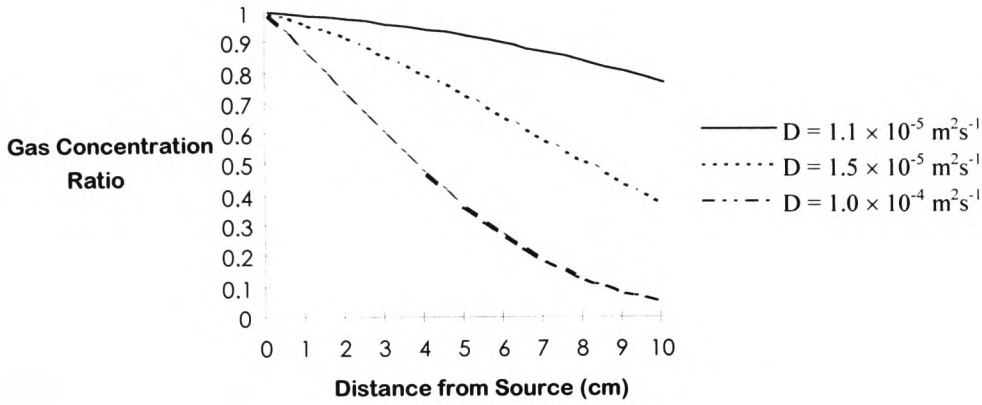
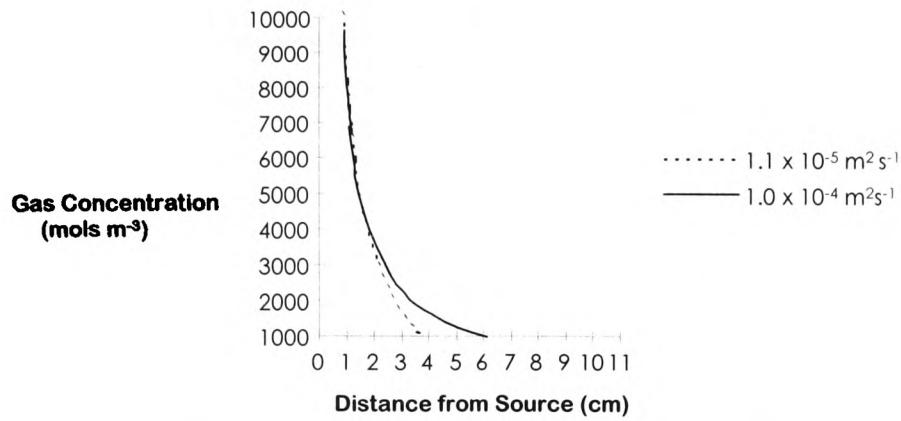


Figure 3.4 illustrates that a complex odour, involving more than one gas, can 'smell' significantly different as distance increases away from the odour source. Therefore, if the odour is sensed close to the source a more representable sample of the analyte gases can be obtained. Figure 3.5 illustrates that odour concentrations reduce sharply as distance increases from the odour source. The distance between the odour source and the sensors is therefore critical since small changes in distance cause large changes in concentration. For example, figure 3.5 shows that gas concentration reduces from 8000 molecules $^{m-3}$ at 1cm to only 2500 molecules m^3. Figures 3.4 and 3.5 also illustrate that concentration variations between different compounds are minimal close to the odour source.

Figure 3.5 Concentration Variation with Distance at Steady State Time



The analysis shows that the concentration of odour molecules, as a result of diffusion, varies greatly depending upon the time after release and also the distance from the source. The human odour sampling system involves numerous odour compounds with differing diffusion characteristics. In a steady state situation the prediction of concentration has been shown to be more simple but the human odour source must be placed at the sensor array interface for a sufficient time to ensure this state is achieved. However, the rate of change of the odour before this steady state period is characteristic of the odour source. Both the dynamic and steady state characteristics can be measured if the odour source remains at the sensor interface for sufficient time.

The concentrations of the odour molecules emitted from a human will also vary according to a number of other factors; for example, bacteria levels on skin, dietary variations and environmental conditions. Compensations can be made for the effects of the environment; for example, the diffusion coefficient can be modified to compensate for temperature and pressure using the following equation[48] :

$$D = D_0 (T / T_0)^{1.75} p_0 / p \qquad (3.7)$$

where

Do is the diffusion coefficient measured at standard temperature and pressure.

To is the standard temperature.

po is the standard pressure.

This equation only compensates for the modification in diffusion in the atmosphere but does not allow for biological changes caused by the environment, which could cause a change in concentration levels. This and other factors affecting the concentration can be viewed as combining, positively and negatively, to modify the number of molecules, *N*, for a particular substance ready for diffusion into the atmosphere. An equation can be developed to show this modification :

$$N = N_0(T / T_0)(r_{h0} / r_h)(B / B_0)\alpha \qquad (3.8)$$

where

N_0 is the concentration at a standardised levels.

r_h is the humidity.

B is the bacteria content of the skin which reacts with secretion to form odour.

α is a compensation factor to account for the variations in measurement systems of the various influencing factors.

The influence of bacteria, *B*, can be illustrated by considering the effect of cleansing the skin. A certain amount of the odour and bacteria is removed when the skin is cleansed which reduces the number of molecules available for diffusion. The bacteria, and consequently the odour, will subsequently increase with time to the level before cleansing.

3.3 An Overview of Mammalian and Electronic Olfactory Systems

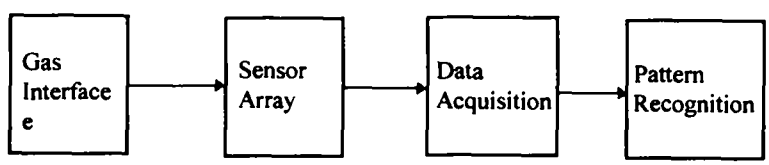
The mammalian nose is a very complex chemical recognition system which is able to discriminate between a great variety of volatile compounds. Unlike sight and hearing, very little is known about the exact mechanisms of mammalian olfaction. The mammalian olfactory system performs the following functions:

- Odour Conditioning : the incoming odour is heated and humidified when it passes over fleshy turbinates [43] which are specialised bones located in the nasal cavity.
- Odour Amplification : mucus covers the actual biological odour sensors and serves the purpose of dissolving the odour molecules which amplifies the odour, sometimes by a factor of a thousand.
- Odour Sensing : the odour then reaches the olfactory epithelium which is coated by a layer of hair-like cilia vastly enlarging the surface area capable of interaction with odours. The odour somehow binds with the receptors on the hair-like cilia which in turn provide excitation to the olfactory sensing neurons to which the cilia are attached.
- Transmission : the olfactory neurons are connected to the brain's olfactory bulb via axons. The olfactory bulb contains bundles of interacting neuron activations, called glomeruli, which decode the resultant odour map from the olfactory epithelium into the odour's essential distinguishing features.
- Perception : the olfactory cortex, helped by various other areas of the brain, finally identifies the particular odour in question and distinguishes the odour from other similar odours which the mammal has encountered.

Electronic olfactory devices, commonly known as electronic noses, are a relatively new technology; the first paper is acknowledged to have been published by Persaud and Dodd in 1982[49]. The definition of an electronic nose is given by Gardner [50] as a device which *'comprises of an array of odorant sensitive*

chemical sensors, each with a non-specific output, coupled to a pattern recognition system'. Figure 3.6 illustrates the general constituents of an electronic nose system. Electronic noses show many similarities to their biological counterparts[50] possibly due to attempts to replicate a tried and tested biological system. The gas interface can be compared to the mammalian nose and provides a suitable environment for the interaction between the gas mixture and the sensor array. The overlapping specificities of the sensor array can be compared to the operation of the cilia in the olfactory epithelium. The data acquisition system converts the sensor variations, as a result of the incoming odour, into an electrical signal. The data acquisition function also amplifies small sensor changes which can be compared to the function of mucus in a mammalian nose. The final constituent of an electronic nose is the pattern recognition system which transforms the multivariate sensor responses into a decision appropriate to the odour in question; in this work an odour must be transformed into a corresponding user. The pattern recognition process can be compared to the olfactory cortex and other related areas of the mammalian brain.

Figure 3.6 Electronic Nose System Diagram



Although there are many ways in which electronic and mammalian noses are related, the electronic version requires a significant period of evolution before a realistic challenge is made to nature. There are, however, many aspects of electronic noses which offer significant advantages over mammalian noses, for example resistance to fatigue and the ability to select sensors for specific applications.

3.4 Evidence for the Uniqueness of Human Odour

The basis for any biometric access control system relies upon the claim that the biometric feature in use is physiologically or behaviourally unique to an individual human being. However, the reputation of a biometric method is attributable to the statistical performance over time; for example, fingerprints have gained their reputation as a result of over one hundred years of usage by criminologists, for the purpose of human recognition.

Since each human is believed to be genetically unique and is therefore unlike any other human, monozygotic identical twins excepted, it is hence assumed that individual biometric features of humans are completely unique. This section evaluates the current evidence which suggests that a human possesses a unique odour which is linked to an individual's genetic make-up. This section also discusses suitable and practical human odour sources for use in an access control system.

3.4.1 Behavioural Evidence

Unlike other members of the animal kingdom, humans do not rely upon their sense of smell for recognition purposes. Many other forms of life rely upon scent in the same way as humans rely upon sight and hearing. Wild cats, for example, use individual scent to map out territories whereas wasps can sense another wasp's pheromonal odour from a great distance away [51]. The characteristic odour of a human does exist even though a human's nose has evolved without the need to recognise a person by odour alone as discrimination by odour plays only a minor role [51].

Dogs are particularly renowned for and adept at recognising a human by use of characteristic odour patterns. Dogs possess an extremely refined sense of smell,

allowing them, for example, to trace the scent of a human which is many days old, with only an article of clothing as a source. Dogs have been used for this purpose for many years in order to track criminals [52]. The smelling power of dogs has also been analysed in a more scientific manner in order to assess their true discriminating capabilities. Dogs have been shown to be successful at distinguishing between the scents of different humans except in the case of monozygotic, identical, twins [51]. This inability of dogs to discriminate between identical twins implies that smell is in fact linked to the genetic make-up of a human.

Dogs have shown the ability to distinguish between the scent from their owners' hands and that of strangers, under scientific conditions [52]. The dogs were trained to retrieve a dumbbell scented by their owners' hands from a series of pairs of dumbbells which had either been scented by other handlers or were unscented. A bloodhound discriminated between its owner's scent and an absence of human scent in 90% of the trials and between its owner's scent and a stranger's scent from 83% of the trials. However, the dogs were unable to repeat these results when the odour was obtained from the armpit of the humans. This result may imply that a unique human odour does not exist but more likely that the dogs were unable to respond spontaneously to this new type of odour source[52].

The characteristic smell of a human is created by the interaction between human secretions and bacteria living on the surface of the skin[53]. Although the armpit is a rich source of odour it is strongly affected by deodorants and anti-perspirants; these chemicals not only suppress the bacterial action of the skin but also emit strong odours which are specifically designed to mask the smell of human odour, hence confusing the dogs.

Similar work has also been performed to investigate the discriminating ability of humans[53,54 ,51]. An experiment involving eight humans[54] concluded that the humans involved were able to discriminate between two individuals by use of the

odour from the palm of the hand only. The experiment was carefully conducted to ensure that the palm odour was not influenced by a strong smelling substance such as lingering smells from onions. Other research has shown that humans are capable of recognising family members by smelling worn clothing[51]. This research concludes that humans are able to discriminate between people to an extent and that individuals do possess a characteristic odour.

This research does not, however, investigate the long term stability of the odour from humans. The odour emitted from a human may be constantly changing but still remain significantly dissimilar from the shifting odour patterns of other humans. Human odour, for example, can be easily modified by contact with strong smelling compounds such as perfumes. The combination of all of these 'external' compounds coupled with the odour resulting from bacterial action on 'naturally' occurring secretion, will produce a highly unique odour pattern for an individual. The odour from a human can therefore be broadly attributed to two main sources;

$$O_{total}(t) = O_{external}(t) + O_{natural}(t) \tag{3.9}$$

The externally influenced odour, $O_{external}$, may be caused by the actual presence of the compound on the skin or indirectly via secretions. The external odour component is constantly changing due to variations in diet, and physical contact with contaminating substances. In an ideal case the electronic odour sensing device must contain only sensors which are not responsive to externally influenced odours. For example, a biometric odour sensing device which is extremely sensitive to a common strong substance such as onions or petrol would inevitably fail to recognise users who have become contaminated. Unfortunately however, current gas sensor technology is not capable of producing sensors of such high selectivity. The overlapping selectivities of the sensors can not, therefore, provide immunity to all external odours, but arrays can be chosen to reduce the interference effect of these odour compounds.

The naturally generated odour, $O_{natural}$, is influenced by a great number of factors. As previously stated, a human's characteristic odour is produced by the interaction of secretions with the microflora of bacteria living on the skin, in the vicinity of the secretion. The majority of secretions are actually sterile and odourless until the onset of bacterial action. The characteristic smell is, therefore, influenced by the content of both the secretion and the bacteria. The content of secretion produced by the body is dependent upon various factors such as diet and hormonal levels. The content of the bacterial microflora has been shown to differ between individuals but some types of bacteria are common to all humans[51]. The odour influencing bacteria has been shown to remain quantitatively stable from day to day[53].

3.4.2 Identification of Human Odour Compounds

There are a variety of different sources which contribute to a human odour, each having a characteristic smell. These odours are usually related to natural effluent products released continuously from a human body. The main sources of these continuous effluent products are oral, in the form of breath and saliva, and from the skin, in the form of sweat and sebum [55]. Although several other effluent sources exist they are intermittent so are not appropriate for a continuous biometric access control system.

The oral odour sources may provide a rich source of odours with high concentrations but also suffer from the effects of contamination and extreme variations in chemical composition. For example, factors influencing contamination of pulmonary gases include diet, health and oral hygiene, air quality, time since last meal, smoking, and physical and psychological exertion[56]. Human breath also consists of high levels of moisture content to which most existing gas sensors are exceptionally sensitive.

Each area on the surface of a human emits a characteristic odour which combines to form the overall odour of a human. Research has been conducted to capture the odours surrounding the surface of a human body [56]. Specially constructed chambers were used to detect 300-400 odour compounds emanating from the skin surface, of which only 135 were identified using Gas Chromatography - Mass Spectroscopy (GC-MS) techniques. Obviously these compounds were a combination of various areas of the body such as the scalp, hands or feet. A sealed chamber within which a human sample must be contained for a fixed period of time is an inappropriate sampling method for a biometric access control device, so a specific area of the body had to be targeted in order to produce a user friendly sampling system.

The skin is covered with various glands which possibly contribute to the unique nature of human odour. The entire surface of the skin is covered with sebaceous glands which continuously produce a liquid called sebum which lubricates the skin. Sebum contains lipid compounds which are a diverse and often unique group of compounds, regarded as responsible for a significant proportion of human body odour[56]. There are two types of sweat glands: apocrine glands, which emit perspiration, and eccrine glands, which produce sweat and regulate body temperature. As discussed previously, the action of bacteria upon this sweat is responsible for human odour characteristics. There are approximately five million eccrine glands in the human body, and about two million of these are in the hands. However, apocrine glands are located primarily in the armpits and groin area which are unsuitable for use in a biometric access control device because of their physical location. The hands, which were previously discussed as unique odour sources, contain high numbers of both sebaceous and eccrine glands which are likely to produce a detectable and uniquely discernible odour.

Previous research on odour compounds only investigated possible links with disease and not with uniqueness. Dr Sommerville et al [57] investigated the

possibility of genetic uniqueness of human body odour. The research used several different methods in order to show the uniqueness of human odour and also to identify the key compounds which are responsible for this uniqueness. Carefully selected humans subjects were chosen for examination and included :

- ten pairs of identical twins of different ages and sexes, three of whom were living apart with different diets,
- several pairs of non-identical twins,
- unrelated volunteers.

Odour samples from all of the volunteers were analysed using gas chromatography. The results showed that the peak pattern variations remained fairly consistent within short time periods. Several regions in the odour spectrums were identified as possessing unique information. Figure 3.7 illustrates a pair of odour spectrums for two unrelated people [58] which indicates many visually distinguishable differences. Figure 3.8 shows odour spectrums for two identical twins [58] which are clearly more visually similar than the unrelated people in figure 3.7. Trials with dogs verified that, using odour spectrum portions of the responses given in figures 3.7 and 3.8, dogs were able to distinguish reliably between unrelated people whilst being unable to distinguish between identical twins[58].

Figure 3.7 Chromatograms from Non Identical Twins

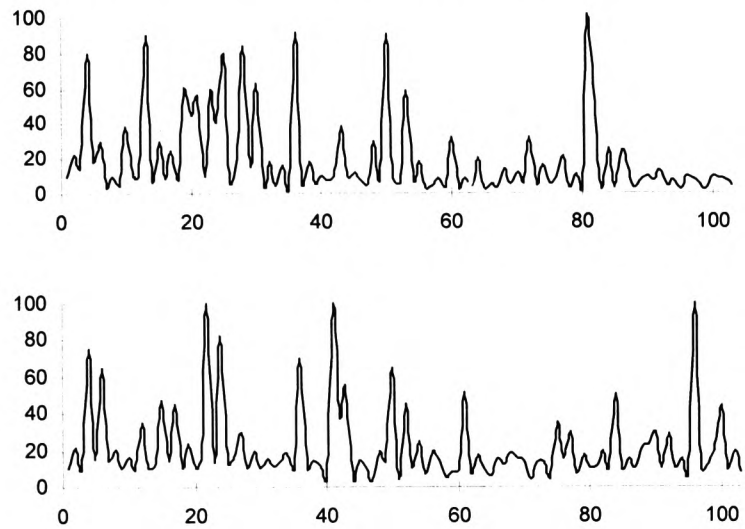
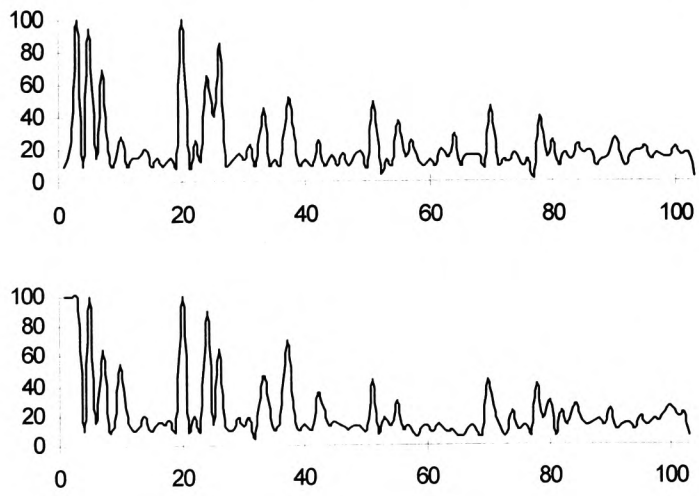


Figure 3.8 Chromatograms from Monozygotic 'Identical' Twins



An attempt was then made to identify the characteristic odour compounds using mass spectrometry. This work showed that the larger peaks were actually common to most humans sampled and were most probably caused by environmental contaminants, basic biological products and common skin-based bacteria acting upon the skin secretions. The compounds responsible for uniqueness were

regarded as being present in small concentrations, some of which were below the resolution of the apparatus used for the test. The research does, however, produce a list of possible compounds which are responsible for a human's unique odour[57].

The conclusion of the research into the area of human odour is one of uncertainty. Although most researchers agree that a unique human odour exists, there is at present only limited theoretical and experimental proof regarding the chemical compounds responsible. The compounds are undoubtedly present in small concentrations and simultaneously intermingled with unwanted contaminants and common human compounds. The implication is that the sensors should possess an extremely finely tuned range and high sensitivity in order to provide a response to the low concentration unique compounds.

This level of specificity is not available at present with current gas sensor technology but does not necessarily imply that the detection of humans by odour is not possible. As previously stated, an array of sensors was needed which provided large overlapping sensitivities to the ranges of gases under scrutiny. The resulting pattern recognition techniques were then used in an attempt to extract the small unique variations from a background of unwanted 'odour noise'.

3.5 Odour Sensors : Electrically Conducting Polymers

Conducting polymers were initially investigated by the U.S. military for use in the Stealth bomber programme in order to aid in the evasion of enemy radar [59]. The University of Warwick investigated the application of conducting polymers to odour sensing which culminated in the first prototype electronic nose in 1982 [59,50].

Electrically conducting polymers have proved to be a good choice for use in an electronic olfactory device for various reasons. The sensors exhibit a reversible change in electrical resistance when exposed to an odour compound which is a variable which can be easily measured by various interfacing methodologies [60]. The sensors show very fast response characteristics at room temperature and also extremely fast recovery times. These two characteristics are especially crucial for a biometric access control device as humans would soon become frustrated if large delays were necessary. Both fast response and recovery are needed as delays during sampling or between samples would severely disrupt an access control strategy. Section 1 discusses the importance of response speed when designing an access control system.

A single conducting polymer sensor responds to a wide range of volatile odour compounds which is advantageous since an array of sensors with overlapping specificities is required. The actual responsivity of a sensor to different odours can be altered by various methods[61]:

- Different monomers, used to create the polymer, produce different response characteristics.
- A monomer can be doped with a substance which alters the odour response characteristics.
- The polymer structure can be altered by varying the way in which polymerisation is carried out.

- The sensor substrate, onto which the polymer is grown, can be modified in order to change odour selectivity.

The methods highlighted above imply that sensors could be 'tailor made' for almost any application by controlling and selecting the appropriate parameters. This flexibility and control are, at present, far from reality since conducting polymer sensor research is still in its infancy, although a great deal of research is being carried out at various institutions around the world. Even though different conducting polymer sensors exhibit differing selectivities they can not currently be produced to respond to a specific odour compound. This is because the way in which conducting polymer sensors interact with odours is not fully understood. The highly flexible control of sensor production also proves to be a disadvantage as the re-producibility of a specific sensor is extremely difficult.

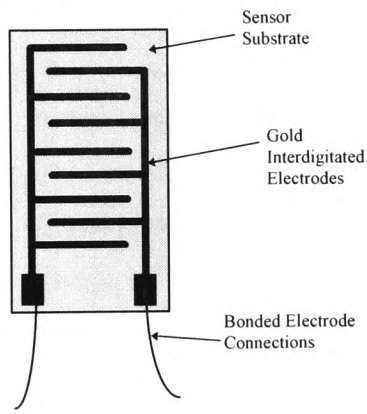
Conducting polymer sensors also have several disadvantages common with most currently available gas sensors, which can be summarised as follows:

- Sensors are sensitive to environmental factors such as humidity and temperature. Common with other sensor types a dependency to environmental effects has been documented[62] although the major effect is a shift in baseline resistance and not a change in the magnitude of response.
- Sensor characteristics change as an effect of ageing and poisoning. However, the effect of ageing has been shown to affect only the baseline resistance and does not affect the magnitude of readings taken from that baseline. Poisoning may be permanent or temporary but measures could be taken to filter these destructive odour compounds before they reach the sensors themselves.
- Sensors cannot be produced to respond to, or reject, a specific odour compound. Only a response to a group of odour compounds with similar characteristics can be obtained.

3.5.1 Sensor Construction

Conducting polymer sensors can be produced chemically or electrochemically both providing sensors of similar characteristics. Electro-polymerisation is usually chosen as greater control of the polymer growth can be obtained. A thin film of the conducting polymer is grown across the narrow gap (< 100μm) between two electrodes. The distance between the electrodes is crucial for the growth of the polymer film. Figure 3.9 shows a typical layout of an inter-digitated conducting polymer substrate.

Figure 3.9 An Inter-digitated Conducting Polymer Sensor Substrate



The electrodes are usually inert gold to avoid any interactions with the chemicals in question. Various substrate materials have been used including ceramic and silicon [63]. The distance between the electrodes is also a critical factor when designing polymer sensors. The electrode gap alters the way in which the polymer grows, the baseline resistance and also the extent to which the sensor resistance changes when subjected to a particular gas. The polymer grows simultaneously from each electrode and eventually joins in between the electrodes to create a film[64]. Since the length of the polymer chains dictate the resistance of the sensors, so the electrode gap is partly responsible for the resistance of the sensors. If the gap is too small then too much polymer will bridge the gap whereas a very

large gap may stop the two polymer growths from meeting, hence creating an open circuit sensor [64].

There are, as previously mentioned, a large number of possible monomers which are possible candidates for electro-polymerisation but the use of several monomers in particular has been well documented. Two such monomers are poly(pyrrole) and poly(thiophene) which are heterocyclic in nature. Electro-polymerisation produces oxidation of the monomer at the electrode surface, which generates a radical cation[65,60]. A dimer is then produced by two cations reacting or alternatively by oxidation of a cation-neutral monomer pair. The dimers then undergo further oxidation and coupling reactions to produce increasingly longer chains which will eventually be deposited onto the electrodes.

The resulting polymer is a polycation which is balanced by an anion from the growth solution. An anion is usually present between every 3 to 6 monomer units in the polymer chain. Consequently the resultant properties of the polymer are also influenced by the choice of counter-ion which is contained in the growth solution, in addition to the influencing factors previously mentioned.

3.5.2 Response Mechanisms

There are many possible mechanisms in which volatile chemicals interact with a conducting polymer to produce a characteristic change in conductivity of the sensor. Although the exact nature of gas adsorption is not known it is believed that the resulting sensor conductivity change is caused by the combined effect of a number of interaction mechanisms [63,65]. Five such contributing mechanisms are [60] :

- Direct charge carrier generation or removal from within the polymer film. This behaviour corresponds to oxidation or reduction of the polymer by the volatile chemical.

- Changes in the mobility of the charge carriers along the polymer chain by interaction of the gas molecules with the polymer itself.
- The odour could interact with the counter ions held within the polymer film. This interaction could cause an increase or decrease in conductivity depending upon the type of gas involved. An electrophilic gas will increase the number of charge carriers resulting in an increase in conductivity whereas other gases may cause interaction at the positive sites in the polymer film resulting in a decrease in conductivity. This mechanism implies that the polymer selectivity and responsivity can be altered by appropriate choice of the counter anion [65].
- The adsorbed odour could modify the intra-chain transfer process resulting in a change in conductivity. The polymer contains many chains which rely upon a chain-hopping process to provide conduction along the entire length of the film. If this chain-hopping process is altered by the interaction of the gas, a conductivity change with then result.
- The presence of the odour at the interface between the polymer and the sensor substrate contacts could affect the rate of charge transfer between the metal contact and the polymer, again resulting in a change in conductivity.

The response of conducting polymers to relevant compounds ranges from seconds to tens of seconds depending upon the sensor and odour compound; the results illustrated in section 4.1.1 confirm this. Conducting polymer sensors also show rapid recovery times in the order of seconds, also confirmed in section 4.1.1. When considering a biometric system, sensors with the minimum response time coupled with rapid recovery times should be chosen. The speeds of response and recovery of this type of sensor are rapid when compared to other types of sensors which exhibit significantly slower response speeds; for example, minutes in the case of tin oxide sensors.

Although these sensors show favourable response and recovery kinetics they do not provide a linear relationship between gas concentration and steady state

response. The non linear form of this relationship can be described by a Langmuir isotherm [63] where the steady state response reaches a saturation value at a certain concentration level. This behaviour can be explained using the aforementioned response mechanisms. The polymer, for example, only possesses a finite number of counter ions which are readily available for interaction with the gas. Once all of these interactions have been performed, the conductivity of the polymer film will remain constant irrespective of any increases in gas concentration. The Langmuir isotherm is shown in figure 3.10 and is described further in section 3.6. The response of the sensors to a wide variety of organic odour compounds characteristic to human odour emissions is also of paramount importance. Different types of conducting polymer sensors have been shown to respond to a range of organic volatile compounds [66].

3.6 Electronic Nose Model

The sensor array can be viewed as converting the sampled 'smell' from odour space to sample space. The odour space can be represented by the concentration level of each of the chemical odour compounds to which the sensors are capable of response. The odour space for a human sample is obviously of high dimensionality as many compounds are released from a human body. Conducting polymer gas sensors do not provide high selectivity to specific compounds so the odour space dimensionality remains high. The sensor array converts the odour into the corresponding sensor space which maps the odour using features extracted from the changes in electrical characteristics measured from the sensors. The ideal sensor array would produce a sensor space which tracks the odour space for the smells which are under consideration. In reality the sensor array is affected by numerous parameters such as low specificity and noise levels which tend to reduce the confidence and reliability of the sensor space[67].

The response from a sensor i in an array of n sensors can be estimated to be [67] :

$$S_i = a_{i0} + a_{i1}x_1^{k_i} + a_{i2}x_2^{k_i} + \ldots + a_{im}x_m^{k_i} \qquad (3.10)$$

where

S_i represents the response of sensor i.

m represents the individual odour compounds in the odour mixture.

x represents the concentration of each of the m odour compounds contained in the measurand odour mixture.

a_i represents a sensitivity coefficient for each of the m compounds.

k_i is an exponent to allow for the non-linearity in sensor response.

The implication is that although a particular sensor is non-specific and responds to a range of compounds the sensor shows higher levels of contribution to particular compounds. The symbol a_{io} represents the output of sensor i in odourless air; this condition is impractical in reality and the factor a_{io} represents the sensor response to the background odour of the environment before the sample odour is introduced.

This response model assumes that the odour compounds are independent from each other and are not chemically affected by the presence of other gases[67]. This assumption is realistic since the majority of odour molecules are polar, see section 3.1 for further details. Interference by unwanted gases is modelled by a non-zero sensor sensitivity, a_{im} , to interference gases, where the interference gases are a subset of the m gases present in the odour sample.

However, this model for response does not account for the various errors in the sensing system. These errors arise from concentration variations of the components in the sample. Concentration variations may be caused by the sampling system or variations in the odour sample itself. The combined effect of all error components can be represented by a fractional error, ε, which contributes

to the response. This error modifies the sensor array equation to the following form [67]:

$$S_i(1+\varepsilon_{S_i}) = a_{i0}(1+\varepsilon_{a_{i0}}) + \sum_{j=1}^{j=m} a_{ij}(1+\varepsilon_{a_{ij}})[x_j(1+\varepsilon_{x_j})]^{k_i} \qquad (3.11)$$

where

ε_{S_i} represents the total effect of error on sensor i expressed as a fraction of the sensor response S_i.

$\varepsilon_{a_{i0}}$ represents the fractional error contribution to the baseline value of sensor i. This baseline error is often expressed as the drift and can be attributed to ageing, poisoning and environmental effects.

$\varepsilon_{a_{ij}}$ represents the fractional error contribution to the sensitivity of sensor i to gas j. This error is attributable to many causes such as environmental effects, variation in the sampling mechanism, sensor ageing and poisoning effects.

ε_{x_j} represents the fractional error contribution caused by natural variations in the odour of the sample.

Since the odour sample in this case is a human, the error component ε_{xj} could be attributable to a variety of different causes as described in section 3.4.1. This error component can only be reduced by attempting to sense odours which are relatively stable; for example, signature odours should be chosen which do not strongly depend upon dietary factors. The reduction of this error is extremely problematic at present because very little is known concerning the exact constituents of human odour and the variability caused by external and internal factors.

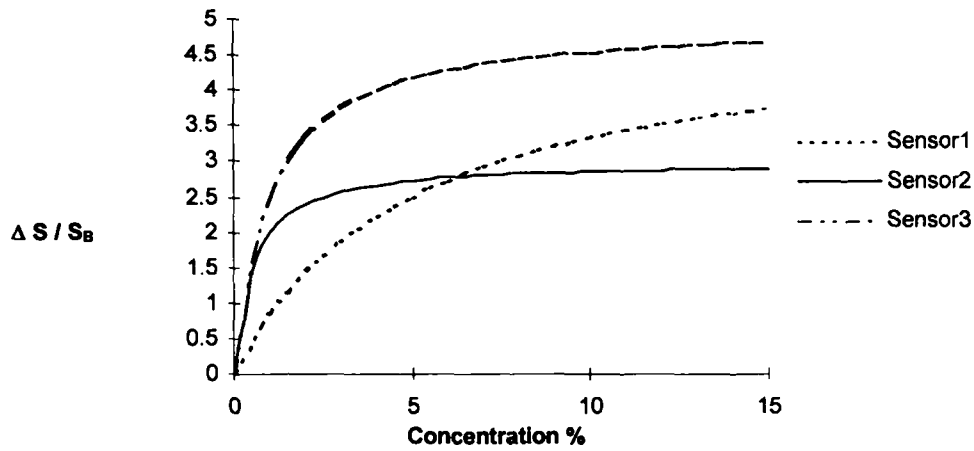
An improved model of the electronic nose accounts more accurately for systematic effects, such as ageing and humidity, and also allows for the possibility

of adsorption saturation. This improved model uses a Langmuir isotherm to allow for saturation and also introduces an error on the exponent k_i to simulate poisoning, ageing or environmental effects more accurately[67,60]. The new model is given by :

$$\begin{aligned} S_i(1+\varepsilon_{S_i}) = {} & a_{i0}(1+\varepsilon_{a_{i0}}) \\ & + \sum_{j=1}^{j=m} a_{ij}(1+\varepsilon_{a_{ij}})[x_j(1+\varepsilon_{xj})]^{k_i(1+\varepsilon_{k_i})} \\ & + \sum_{j=1}^{j=m} \frac{b_{ij}(1+\varepsilon_{a_{ij}})x_j(1+\varepsilon_{xj})c_{ij}(1+\varepsilon_{c_{ij}})}{1+b_{ij}(1+\varepsilon_{b_{ij}})x_j(1+\varepsilon_{x_j})} \end{aligned} \qquad (3.12)$$

Figure 3.10 demonstrates the predicted variation in steady state response for ethyl alcohol for three different sensors. The sensor response shown in figure 3.10 is $\Delta S / S_B$ which represents the change of response, ΔS, divided by the baseline response at 0% concentration, S_B. This graph shows that sensors do not have the capability to increase response, for example conductivity, consistently as concentration increases, the possible reasons for this attribute were discussed in section 3.5.2.

Figure 3.10 Steady State Response as a Function of Odour Concentration for Three Different Sensors



3.7 Determination of Array Size

The concept of using sensors with overlapping sensitivities has been previously discussed in section 3.3. The actual number of sensors required for an array is dependent upon many factors such as the variation of sensor response and environmental effects. However, the array size can be estimated by analysing both the performance capabilities of the sensors and also of the data acquisition system [68].

3.7.1 Data Acquisition Capabilities

If the odour sensing array contains p elements and the Analogue to Digital Converter (ADC) quantises the sensor signal to n bits then the highest possible number of patterns which can be generated by the array is given by N [68],

$$N = 2^{pn} \tag{3.13}$$

It is evident that even a very small array of sensors is capable of generating vast numbers of different patterns. A two element array with a 8-bit ADC, for example, is capable of generating 65536 different patterns. This degree of resolution is, however, misleading because the sensors would be required to respond non-identically to the odours presented to the array. This accuracy is not feasible in a practical system so arrays incorporate a degree of redundancy to compensate for the reduced discriminating power of the system. Various factors can be shown to reduce the resolution of the ADC which in turns reduces the discriminating capabilities of the sensor array [69].

3.7.2 Sensor Response Variation

The response from a sensor is dependent upon many factors such as humidity, distance from the odour source, ageing effects of the sensor and consistency of the odour sample itself. These factors increase the variation in response of the sensors which effectively means the sensor does not produce an identical response each time a specific odour is sampled. Figure 3.11 shows the effective reduction in available quantisation levels when the level of sensor response variation is expressed as the percentage of the ADC range which results from samples of the same odour[67].

Figure 3.11 Effect of Sensor Variation on ADC Resolution

% Sensor Response Variation	Effective Number of Quantisation levels		
	n=4	n=8	n=12
0.01	16	256	4096
0.1	16	256	1001
1	16	101	101
10	11	11	11
20	6	6	6

Figure 3.11 illustrates that if sensor response variation to the sample odour compound is high, for example 20%, then even when a 12 bit ADC is used the number of levels to which the sensor response can be assigned is only 6.

Therefore, high response variation limits the resolving power of the odour sensing system.

3.7.3 Probability of Pattern Generation

A finite probability can be calculated that given a number of different odours, the array will generate an original pattern[68]. The probability of a sensor signal having a magnitude s grey level is

$$P(s) = \frac{1}{g} \tag{3.14}$$

where

$g = 2^n$ for an n-bit ADC. (3.15)

This result can be used to determine the probability that a single sensor will produce original patterns for each odour, m, presented to the device[68].

$$P(s) = \frac{g!}{(g-m)!g^m} \tag{3.16}$$

This expression can again be modified to account for an array of sensors, of dimension p, instead of a single sensor in isolation[68].

$$P(s) = \frac{g^p!}{(g^p-m)!g^{pm}} \tag{3.17}$$

and substituting g for an n-bit converter

$$P(s) = \frac{2^{np}!}{(2^{np}-m)!2^{npm}} \tag{3.18}$$

This equation provides the estimated probability that an array of p sensors, using an n bit ADC, will produce a different and discernible multivariate response when exposed to m different odour patterns.

To specify an initial sensor array size for the human odour sensing device, a number of performance parameters must be determined. The number of odour patterns m must be estimated which is equivalent to the number of people who will use the system. This number must include the number of people enrolled onto the system in addition to an allowance for impostors. This biometric system is intended for a small field trial involving a maximum of fifteen people so the number of unique odour patterns can be estimated at 30 which also includes an estimated allowance of fifteen odour patterns for intruder attempts. The probability of a unique pattern, $P(s)$, can be set to the average recognition rate obtainable by a commercial biometric access control systems[19] which is approximately 0.999, meaning 99.9% of people can be recognised correctly. Using the equation above, if the full range of the ADC is utilised then the number of patterns required to generate original odour patterns for each person is approximately 4.5×10^{16}. This can be fulfilled by using an 8-bit ADC and seven sensors.

However, as previously mentioned, the effective resolution of the data acquisition system is severely limited by a number of sources. These sources must be accounted for when specifying the sensor array size.

- Human - Sensor Interface : as described in section 3.8.2.1, the sampling grid introduces an estimated response variation of 5%.
- Sensor Baseline Drift : as described in section 4.1.2 the variation in baseline, due to the effects of temperature and humidity, caused a reduction in the ADC range of 8%.

- Sensor Response Variations : the average response variation to a set of odour compounds, over a period of five days, for the conducting polymer sensors utilised has been found to be approximately 5% [70].

Hence the total estimated sensor response variation is approximately 20%. This variation has a significant effect upon the effective resolution of the data acquisition system. Referring to figure 3.11 it can be demonstrated that the effective number of quantisation levels has been reduced from 256 to just 6. The estimated array size is now 16 sensors as opposed to 7.

3.8 Human Odour Sampling

Using the design criteria which has been identified, a prototype odour-based access control system was constructed. This prototype was used in a field trial to gather data from a group of people on a daily basis. Various problems associated with the prototype are described in this section. These problems can divided into several distinct areas:

- Optimisation of the Sensor Array for Human Odour.
- Design of the interface between a Human and the 'Electronic Nose'

3.8.1 Sensor Selection

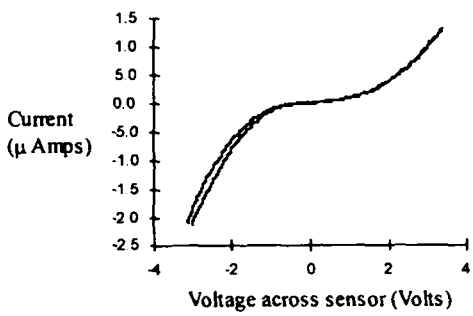
Sensors were selected for the biometric system by subjection to a number of tests aimed at producing an optimal sensor array. The tests were performed in the following sequence :

- Electrical Characteristics : The baseline resistance of the sensor was assessed by producing the Voltage-Current (VI) characteristics and figure 3.12 illustrates the resulting VI curves for typical conducting polymer gas sensors. Sensors were chosen demonstrating resistance in the range 100KΩ to 20MΩ at an operating voltage level of 3.5 Volts d.c, this ensured compatibility with the

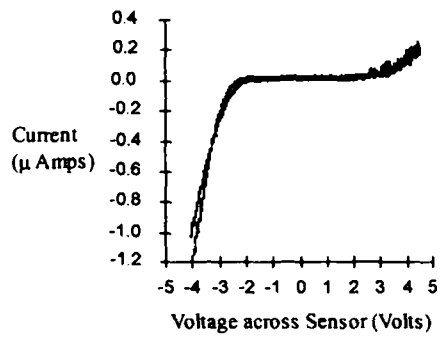
d.c. sensor addressing technique employed. Figure 3.12a illustrates a stable sensor with an acceptable baseline resistance at most non-zero DC voltage levels whereas the sensor shown in figure 3.12b shows an unstable sensor which appears almost open circuit between -3 to +3 Volts. The unstable sensor exhibited a rapidly decreasing resistance when the voltage was increased negatively, whereas when the voltage was increased positively a slight increase in resistance was observed which was unstable. Sensors showing such instabilities were avoided for use in this prototype instrument. However, the sensors may well show greater stability when used with a different addressing technique, such as alternating current or multi-frequency.

Figure 3.12 VI Characteristics of a Conducting Polymer Gas Sensor

a. Stable sensor



b. Unstable sensor



- Sensors were subjected to concentrations of thirty seven synthesised gases. The tests were performed at the University of Leeds who were responsible for the manufacture of the sensors [70]. Fifteen of the thirty seven gases were thought to contribute to the unique nature of human odour [58]. The level of concentrations were higher than those emitted from a typical human but were used to evaluate the response of the sensors to known constituents of human odour. Figure 3.13a [70] shows an acceptable sensor demonstrating good responses to the human odour compounds whereas figure 3.13b [70] shows a rejected sensor which responds only to several gases which were not unique human odour compounds.

Figure 3.13a Accepted Sensor Response to Screening Odour Compounds

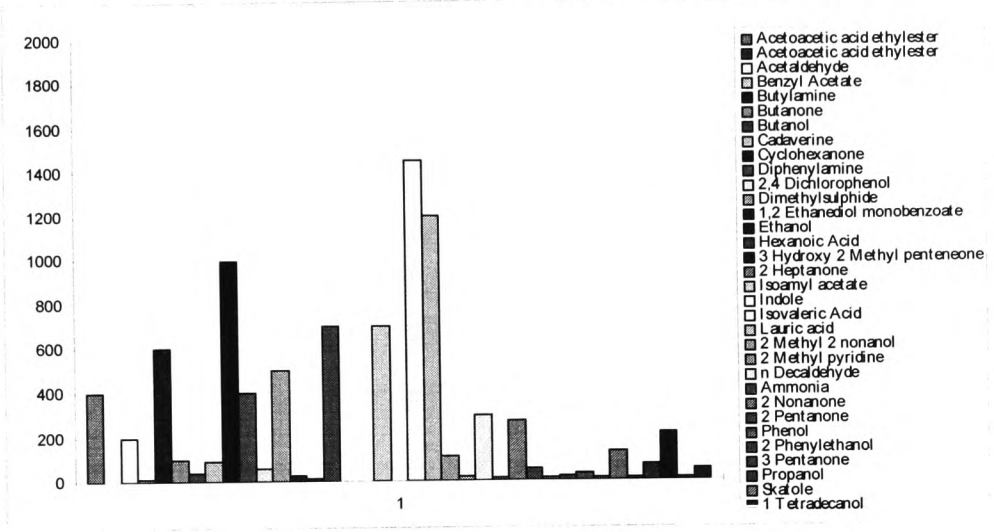
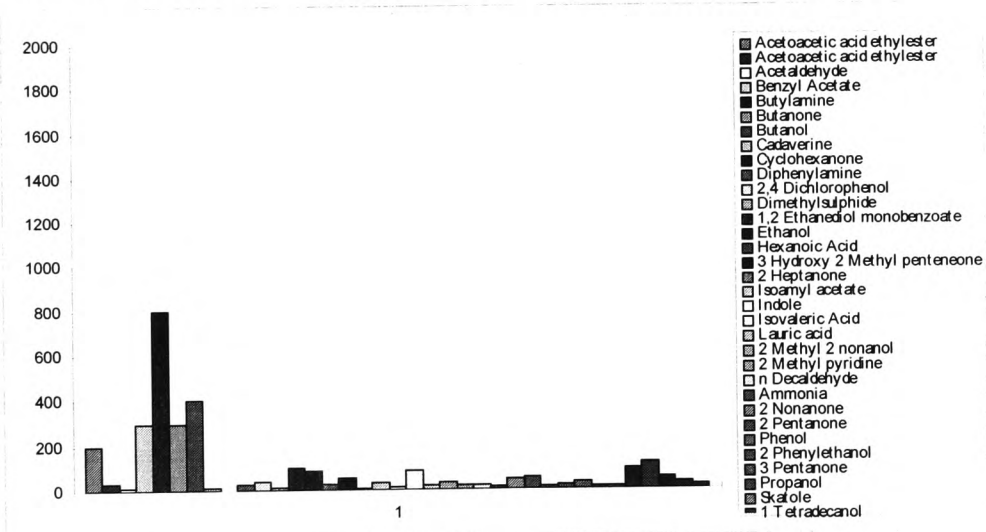


Figure 3.13b Rejected Sensor Response to Screening Odour Compounds



- The adsorption time constant was determined by measuring the time taken for the signal magnitude to reach 90% of the final value and the desorption time constant was the time taken for the sensor resistance to reduce by 90% from the final value[71]. Sensors were chosen which exhibited adsorption and desorption time constants of less than 15 seconds. Several sensors showed

tolerable adsorption characteristics but slow desorption characteristics, sometimes requiring several hours to return to baseline values.

Following trials upon 30 sensors, a set of 21 potentially suitable sensors was compiled. The final reduction of the sensor set was performed by choosing sensors of differing polymers or sensors with the same polymer but differing polymerisation methods. This produced an array of sensors which responded to different, but overlapping, groups of odour compounds since the specificity of a conducting polymer sensor is partially dependent upon the monomer and the method of polymerisation, as described in section 3.5.1.

3.8.2 Human Odour - Electronic Nose Interface

There are two main methods of sensing the gases emitted from an odour source: static and dynamic sampling. Both methods were assessed in order to produce the most effective sampling system within the constraints presented by the sensors and the complex nature of a human odour source. Well established techniques exist for taking solid (biopsy) and fluid(blood and urine) samples from humans but procedures for taking odour samples from humans have not yet been developed [70]. Suggested procedures include a 'whole body collection system' requiring several hours of sampling [55] and collection of under arm axilla sweat using lint pads for a period of 24 hours[57]. These two methods are impractical for a biometric system due to time scales and inconvenience to the user. The odour sampling system had to conform to the following requirements[70] :

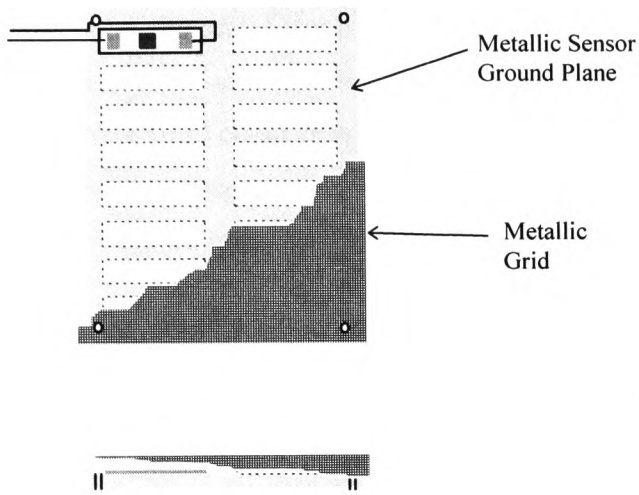
- A complete sample had to occur within one minute; this one minute interval included baseline measurements, odour sampling and sensor recovery.
- A non-invasive technique had to be employed which was comfortable and easy to use.
- The odour source had to contain a sufficiently high odour concentration so that a response could be observed within the short sampling period.

3.8.2.1 Static Sampling Method

Static sampling relies upon the diffusion characteristics of the gas to create an interaction at the sensor surface and requires either a self contained gas sensing chamber where the gas is not able to escape during testing or an interface which relies upon close proximity between odour source and sensor. This close proximity is needed in an unbounded environment since the gas is free to diffuse into the atmosphere. This method, therefore, relies upon a static environment devoid of any gas flow.

The possible sources of human odour were discussed in section 3.4.2 which indicated that hands were a ready source of distinguishable odour representing a discrete, socially acceptable and easily accessed source. The hand is suited for the static sampling method since it can be easily placed onto a sampling grid mechanism. The sensors required very close proximity, within 1cm, of the hand to provide a detectable response, implying odour concentrations were low. The electronic gain levels were increased so that this close proximity could be reduced but proved to be unsuccessful due to excessive baseline drift which reduced the dynamic range of the data acquisition system and caused unreliable sensor readings. This close proximity corresponded to the region of initial relatively high concentration, found near the odour source. This observation corresponded to the odour concentration predictions which illustrated that odour concentration rapidly reduces as the distance from the source increases. Close proximity of the sensors to the odour source, the hand, has the benefit that the source odour is well represented when the distance from the source is less than 1cm. As distance increases from the source, the lower molecular weight molecules are more abundant then heavier molecules so the odour 'smells' differently as distance increases, as previously discussed in section3.2.

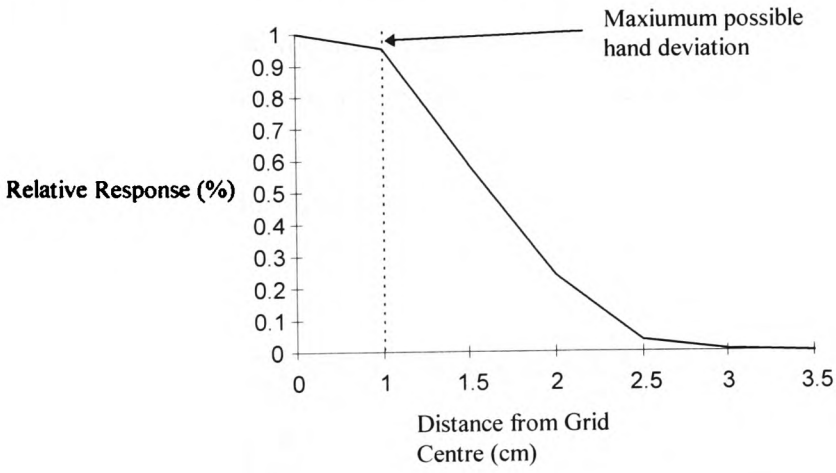
Figure 3.14 Static Sampling System for Sensing Human Odour



The sixteen sensor array was mounted upon a grounded metal backplane, to reduce electrical noise, in a compact pattern in order to reduce the total surface area of the array, as illustrated in figure 3.14. The surface area of the array was critical since ideally the entire array had to be covered by the hand, irrespective of the actual size of the hand. A metallic grid was constructed over the sensor array at a distance of 80mm. Consequently the sensors were exposed to the external environmental conditions which may have minimised environmental imbalances between the sensors and the odour source. The effect of contamination of the grid was minimal since the sensors returned to the original baseline level following an exposure to hand odour. The sampling grid was covered by a shield which reduced air flow to the sensor during sampling which could have caused desorption and consequently inaccurate results. This shield also aided in the locating of the hand and reduced sampling error by restricting horizontal hand movement. The grid was also recessed into its mounting box to aid positioning of the hand and again to reduce air flow over the sensors which could cause desorption. The hand also prevented air flow to the sensors by temporarily blocking the air flow.

The sampling error of the grid was assessed by measuring the average response from the array as a hand was progressively moved horizontally away from a central position on the grid. Figure 3.15 demonstrates that response variation was less than 5% until the hand was moved 1cm horizontally. This distance corresponds to the hand odour becoming out of reach to half of the sensors, since the sensors were arranged in two vertical banks of eight sensors. Vertical movement of the hand had a much lower effect because the response of pairs of sensors reduced as the hand was moved vertically. Consequently the shield was arranged to reduce horizontal movement for an average sized hand, to 1cm from the centre line. This corresponded to a maximum sampling error of approximately 5%. Despite the previously mentioned disadvantages this method consistently produced samples of high magnitude.

Figure 3.15 Effect of Hand Position upon Array Response



3.8.2.2 Dynamic Sampling Method

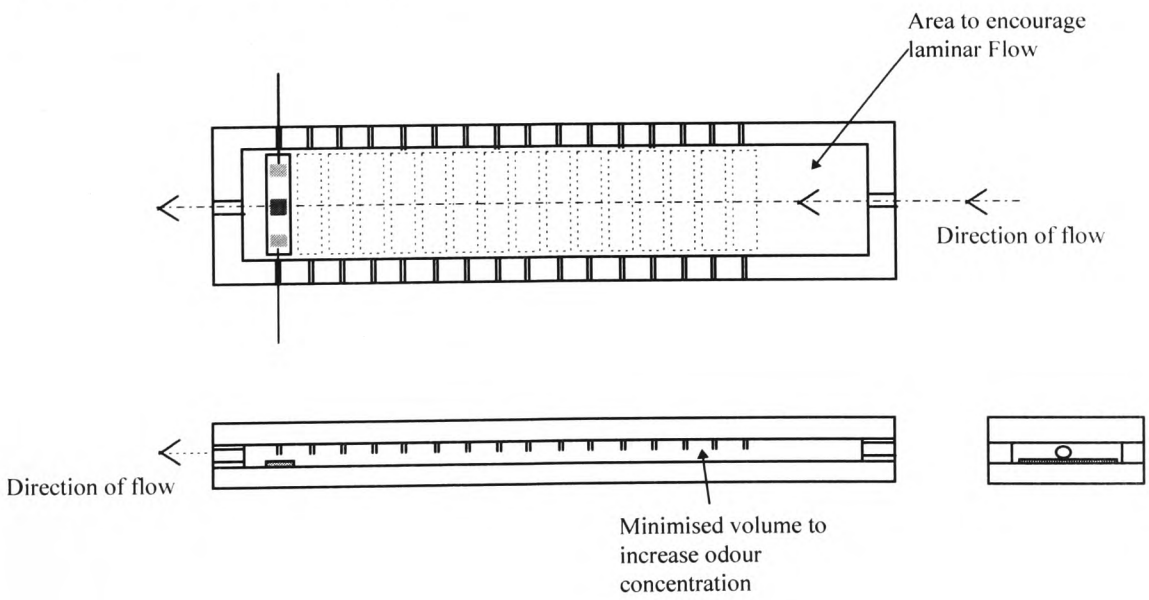
Dynamic measurement techniques use a carrier gas which also contains the measurand gas, human odour in this case. The measurand odour may be extracted from the sample by means of a suction device. The odour flow is then directed to the sensors usually by incorporating a flow chamber. This method enables the temperature and humidity of the sample gas mixture to be controlled reducing the

effect of these influencing factors. There are several disadvantages of this method such as reduced concentration of measurand gases[70], inability to predict gas concentration using diffusion modelling and design problems ensuring adequate and equal coverage to all sensors in the flow chamber.

This method can utilise any area of skin for odour sampling since only a flow of gases is needed to draw the odour from the skin into the flow chamber. However, different areas of the human body possess different concentrations of glands and hence body odour. The hand was again utilised in these trials because of its abundance of odour and its ease of use.

A flow chamber was constructed in order to assess this method of sampling and is illustrated in figure 3.16. Care was taken to reduce the effect of turbulence, which could cause inconsistency and is unpredictable, by incorporating an 'empty' area before the sensors to encourage laminar flow of the gas. The volume of the chamber was reduced forcing the gas into contact with the sensors as opposed to flowing above the sensors, and to prevent excess dilution of the analyte gases.

Figure 3.16 Flow Chamber Design



Initial trials of the flow chamber design utilised an open ended tube which was held in the hand of the person under test. The resulting responses from the sensors were of low magnitude owing to the reduced concentration of the gas flow. The flow rate was systematically varied in order to deduce the optimum flow rate through the chamber. However, this adjustment rate did not increase the sensor responses to an acceptable level. Test samples did not generate an acceptable response either even though they were of a higher concentration than human odour emitted by skin on hand. Use of this combination of sensors, electronics and flow chamber proved to be unsuitable for detecting the human odour compounds concerned in this work.

4. Data and Feature Analysis

A field trial was conducted which covered a period of six weeks with sampling occurring each day during the working week, Monday to Friday inclusive. Twelve volunteers contributed to the trial with members of both sexes and various ages being represented. A minimum of ten samples per day were requested from each volunteer which was usually provided during one session due to work commitments.

The trial resulted in a total of 1782 samples which were provided by eight reliable volunteers who were sampled on a minimum of 93% of the trial days, whilst two volunteers provided samples on 100% of the trial days. The sample data was not treated in any way and was a digitised representation of the sensor array responses.

A number of experiments were undertaken on the field trial data in order to verify the predicted characteristics and relationships within the data. The experiments were divided into two sections :

- data analysis
- feature analysis

Data Analysis involved statistical tests in order to investigate a number of relationships, for example humidity dependency, and prepared the data for feature analysis and pattern recognition.

Feature analysis derived and analysed characteristic attributes from the sensor data. Feature selection determined the feature subset which provided the greatest discrimination between humans. The actual discrimination contribution of the individual sensors was assessed to avoid information duplication and redundancy. Following this, feature extraction techniques were applied to the data in order to assess the discrimination capability of the electronic nose.

4.1 Data Analysis

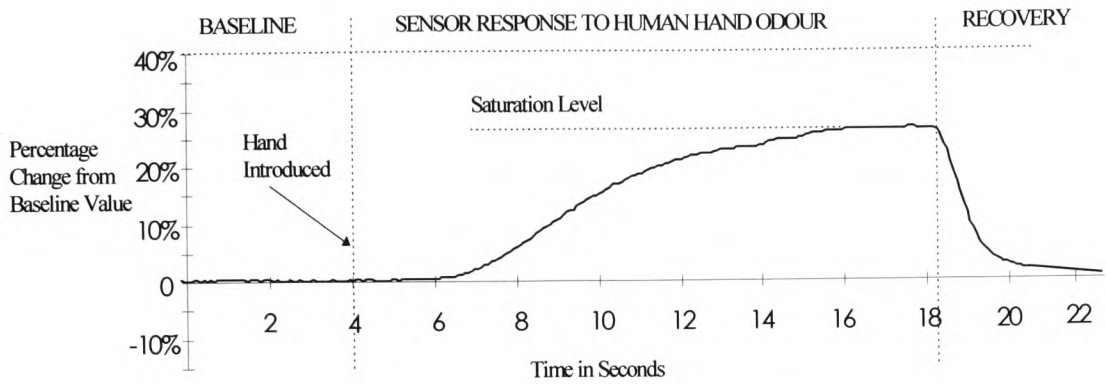
4.1.1 Sensor Response Kinetics

A typical sensor response when exposed to human odour was as shown in figure 4.1. The initial portion of the response, called the baseline, was a measurement of the background odour, taken in the absence of the human sample. The average of the baseline was used to calculate the subsequent features which described the sensor response in a more compact manner; the baseline was therefore the sample reference. A total of 50 samples were measured from each sensor in this region; the sampling rate for each sensor was 60mS which resulted in a baseline period of 3 seconds.

The human odour stimulus was introduced following the baseline readings. This response region can be divided into two main regions: the adsorption region and the saturation region. The adsorption region contains unique information regarding the odour source[73]. The adsorption gradient was dependent upon both the diffusion characteristics of the odour travelling from the hand and also the response mechanisms of the sensors. A total of 250 samples were measured from each sensor in this region, the sampling rate was 60mS which resulted in a adsorption / saturation period of 15 seconds.

The final region of the response curve shows the recovery characteristics of the sensor when the odour sample was removed from the sensor. This region is termed the desorption region, signifying the release of odour molecules from the polymer surface. This region also contains unique information in a similar manner to the adsorption region. The gradient of these regions indicates both the type of odour molecules adsorbed and the extent of adsorption into the polymer structure[73]. A total of 70 samples were measured from each sensor in this region; the sampling rate was 60mS which resulted in a desorption period of 4.2 seconds.

Figure 4.1 Sensor Response Kinetics



The sensor curve shown in figure 4.1 represents an ideal response with the saturation level reached within the time allocated to the adsorption period and also the recovery returning to baseline before the end of the allocated recovery period. In reality, these two criteria were not fulfilled each time a sample was made. For example, one user caused an extremely rapid response whereas a different user only produced a slow drift without reaching the saturation level. Conversely the same may have applied to the desorption gradients. These variations in response were not deficiencies but indicators of individuality which were utilised to increase the discriminating capabilities of the system. However, some sensors did exhibit slow response kinetics or more frequently slow recovery times, with some sensors requiring hours to return to baseline values.

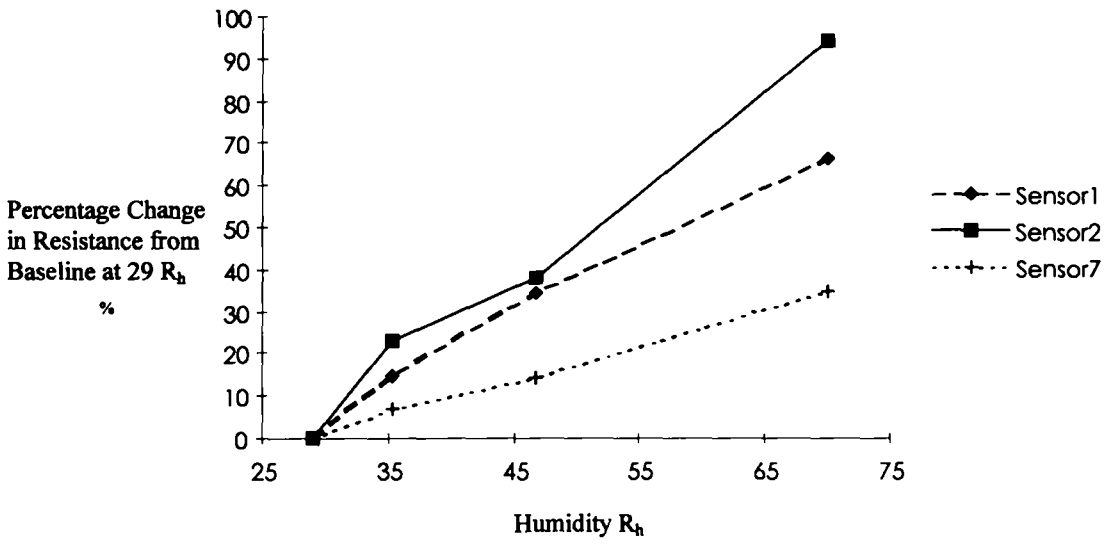
4.1.2 Effect of The Environment

The characteristics of most commercially available sensors are reliant upon the environment in which they operate. The effect of temperature and humidity upon the array of conducting polymer gas sensors was assessed. The array of sensors was placed into an environmental chamber so that sensor resistances could be measured as temperature and humidity were varied [74]. From the results obtained it could be seen that although temperature had a linear effect upon the baseline resistance of the sensor it appeared to have little effect upon the response measured from the shifting baseline level. The percentage change in resistance due to temperature changes was seen to be less than 0.2% per degree Celsius. Hence since

measurements were made using the baseline as a reference point, it follows that the effect of temperature could be ignored.

The effect of humidity produced similar results to temperature except that much larger changes in baseline resistance were observed, as illustrated in figure 4.2 [74]. The minimum and maximum humidities recorded during the field trial were 28 Rh and 45 Rh respectively, which resulted in a maximum change in baseline resistance of 8%, as measured by the sensor addressing technique detailed in appendix 1. This shift in baseline did not affect the magnitude of response since readings were taken from the baseline level. However, this change in baseline reduced the effective dynamic range of the data acquisition system and from this data it can be inferred that information could be lost if humidity levels were more extreme.

Figure 4.2 Variation of Absolute Baseline Resistance with Humidity



The effects of temperature and humidity are dependent upon the sensor under consideration; for example figure 4.2 demonstrates that the baseline resistance of Sensor1 deviates by nearly 100% throughout the humidity range whereas Sensor7 only deviates by 27%. The effects of temperature and humidity can be ignored as long as the readings are taken from the baseline level and the shift in baseline does not shift the sensor out of the range of the data acquisition system. However, if an odour sample exhibits temperature or humidity which is higher than the sensor

array environment then a dis-proportionate response may be produced due to combination of the actual response to odour and a shift in resistance due to the temperature and humidity of the source. For example, a human who has recently exercised would possess increased skin temperature and humidity levels which could mask the measurand odour levels. However, from the results obtained no correlation could be made to response levels and the humidity levels measured for each sample, which indicates that the variation in response levels were mainly caused by variations in human odour levels or sensor response mechanisms.

4.1.3 The Feature Set

Much of the previous work involving gas sensor analysis utilised one feature per sensor; the maximum peak response. Using only the maximum peak response discards the potentially information rich adsorption and desorption areas as well as other potentially useful features.

The data was preprocessed before any feature calculations were made so that sample readings in the adsorption, saturation and desorption regions were calculated as offsets from the average baseline region. This process reduced the effect of sensor baseline drift which was caused by several factors such as humidity, and can be represented in Ohms as :

$$R_n = \left(S_{nB+1} - Bm_n\right) + \left(S_{nB+2} - Bm_n\right) + \ldots + \left(S_{nB+A} - Bm_n\right) \qquad (4.1)$$

where

S is the un-processed sensor response produced by the data acquisition system in Ohms, as defined in equation 3.10.

n is the sensor number.

A is the number of samples in the adsorption, saturation and desorption regions.

Bm is the mean of the baseline resistance readings defined as :

$$Bm_n = \frac{1}{B}\sum_{i=1}^{B} S_{ni} \tag{4.2}$$

where

B is the number of samples in the baseline region.

Features 0 to 10 were calculated directly from this baseline referenced data and can be represented as :

$$F_n = f_0(R_n) + f_1(R_n) + \ldots + f_C(R_n) \tag{4.3}$$

where

f is the appropriate feature function.

C is the number of features.

n is the sensor number.

The resultant features were then normalised across the entire array which attempts to reduce experimental errors in the odour concentrations [75,73] caused by : temperature / humidity dependency, odour interference, sampling error and changes in sensor response mechanisms. The normalisation process can be defined as[73] :

$$FN_{ni} = \frac{F_{ni}}{\sum_{m=1}^{N} |F_{nm}|} \tag{4.4}$$

where

n is the sensor number.

i is the feature number.

N is the number of Sensors.

FN is the separately normalised features.

Features 11 to 15 were calculated from a standardised sensor response data set. The data set was standardised to Z-scores by subtracting the sensor response mean from an individual sensor reading and dividing by the sensors response standard deviation, defined as [76]:

$$Z_n = \frac{R_n - \mu_n}{\sigma_n} \tag{4.5}$$

where

n is the sensor number from an array of A sensors.

Z is the Z-score standardised data set.

μ is the mean of the sampled data.

σ is the standard deviation of the sampled data.

Using Z scores removes all effects of offset and measurement scale. For example, two sets of measurements taken using different units could be compared directly after using Z-scores. Scaling to a Z-score reduces the measurements to comparable units.

The features 11 to 15 can now be represented by :

$$F_n = f_{11}(Z_n) + f_{12}(Z_n) + \ldots + f_{11+C}(Z_n) \tag{4.6}$$

where

f is the appropriate feature function.

C is the number of features.

The calculated features 11 to 15 were then linearly normalised separately in the range 0 to 1 so direct comparison could be made with features 0 to 10.

The sixteen feature calculation functions are illustrated in figure 4.3 and can be defined as :

0. Normalised mean value of the sample.

$$F_{n0} = \frac{1}{A}\sum_{A} R_n \tag{4.7}$$

where

A is the number of samples and n is the number of sensors.

1. Normalised time period to reach the mean value.

$$F_{n1} = Time\ (F_{n0}) \tag{4.8}$$

2. Normalised standard deviation of sample.

$$F_{n3} = \sqrt{\frac{1}{A}\sum_{A}(R_n - F_{n1})^2} \tag{4.9}$$

3. Normalised maximum divergence of the sample.

$$F_{n3} = Max\ (R_n) \tag{4.10}$$

4. Normalised time period to reach maximum value.

$$F_{n4} = Time\ (F_{n3}) \tag{4.11}$$

5. Normalised maximum adsorption gradient.

$$F_{n5} = Max\left(\frac{dR_n}{dt}\right) \tag{4.12}$$

6. Normalised time period to maximum adsorption gradient.

$$F_{n6} = Time\ (F_{n5}) \tag{4.13}$$

7. Normalised maximum desorption gradient.

$$F_{n7} = Min\left(\frac{dR_n}{dt}\right) \tag{4.14}$$

8. Normalised time period to maximum desorption gradient.

$$F_{n8} = Time\ (F_{n7}) \tag{4.15}$$

9. Normalised area contained under response curve.

$$F_{n9} = \int R_n dt \tag{4.16}$$

10. Normalised adsorption - desorption ratio.

$$F_{n10} = \frac{F_{n5}}{F_{n7}} \tag{4.17}$$

11. Maximum divergence of standardised sample.

$$F_{n11} = Max\ (Z_n) \tag{4.18}$$

12. Maximum adsorption gradient of standardised sample.

$$F_{n12} = Max\left(\frac{dZ_n}{dt}\right) \tag{4.19}$$

13. Maximum desorption gradient of standardised sample.

$$F_{n13} = Min\left(\frac{dZ_n}{dt}\right) \tag{4.20}$$

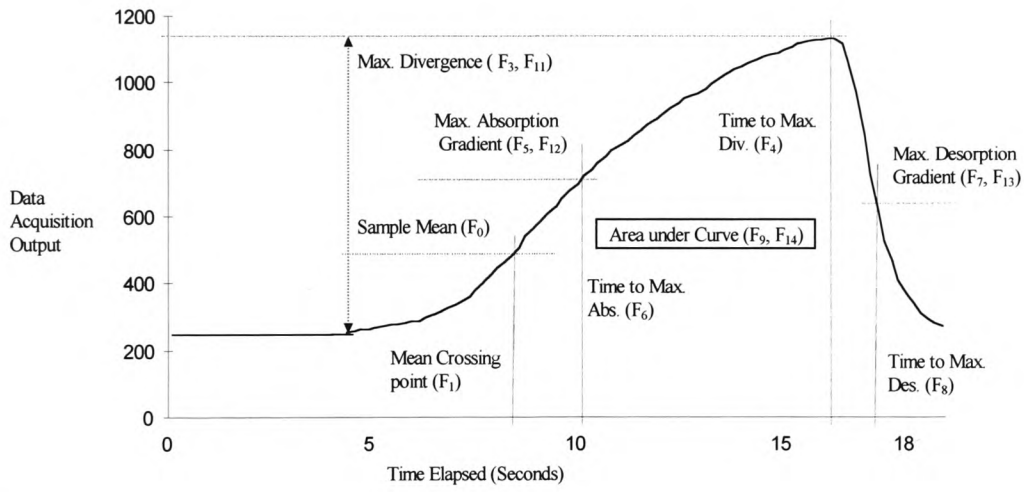
14. Area under response curve of standardised sample.

$$F_{n14} = \int Z_n dt \tag{4.21}$$

15. Adsorption - desorption ratio of standardised sample.

$$F_{n15} = \frac{f_{n12}}{f_{n13}} \tag{4.22}$$

Figure 4.3 Features Derived for Each Sensor Response



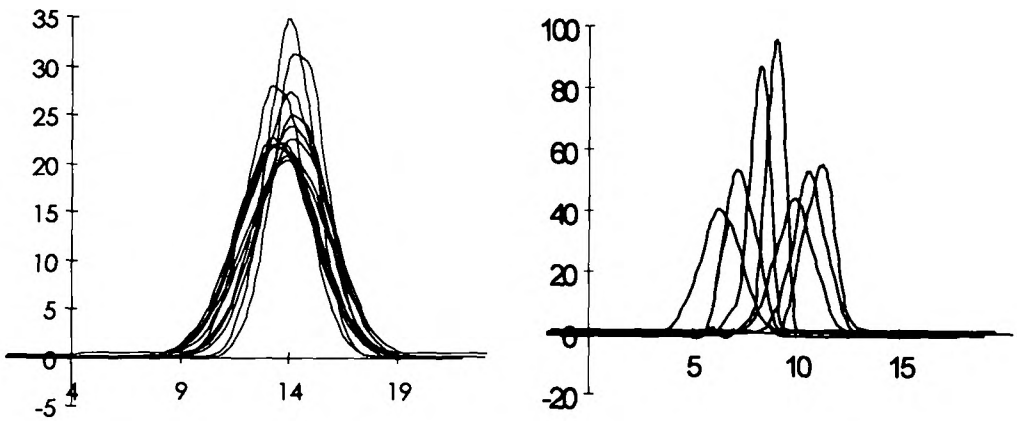
4.1.4 Feature Distribution

The distribution of features for each different class, or person, was analysed to ascertain the degree of interference between classes for each sensor. The interference between two classes can be described as the probability of an observation from class A belonging to class B. The probability density function (pdf) of each class for a responsive sensor is demonstrated in figure 4.4a and was constructed using the maximum divergence feature and assumed a normal distribution. When all class responses were compared together for this responsive sensor it was evident that the sensor was responding in a similar manner for all classes. This common response was most probably caused by an unsuitable choice

of monomer and polymerisation protocol which produced a sensor which responded mainly to a common human odour compound as opposed to genetically unique compounds.

The pdf illustrated in figure 4.4b show the increase in separability between classes for a superior sensor choice. It can be noted that even though the sensor provides superior discrimination characteristics the classes still exhibit overlapping distributions, which would cause confusion when attempting to make a classification decision. When all of the chosen features are used and a number of sensors are considered a more informed decision can be made by combining the class membership probabilities for each individual sensor feature.

Figure 4.4 a. Class Distribution Sensor x b. Class Distribution Sensor y
Based upon Maximum Response and a Normal Distribution within each class.



4.1.5 Statistical Outlier Detection

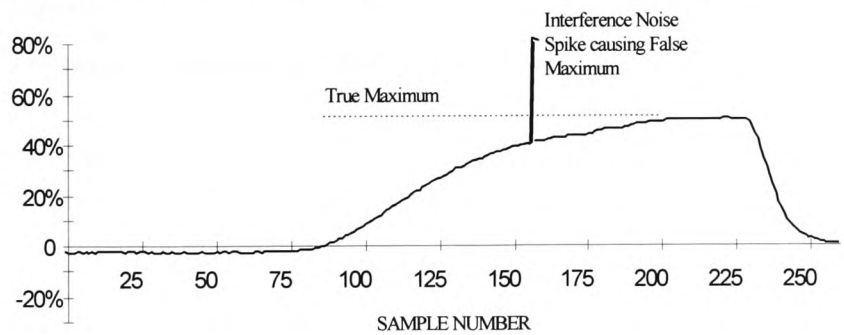
Checks were performed for statistical outliers before commencing with feature selection and future pattern recognition. These outliers may have been introduced into the data in many different ways but can be defined as a variable with a value which is regarded as significantly beyond the natural boundaries of the data [77]. Outliers may be destructive in nature or could indicate distinctive trends in the data. The nature of each specific outlier was considered separately in order determine whether the outlier should be removed, corrected or remain in the data.

There are three different types of outlier which could have existed in the data collected from the sensor array; univariate, bivariate and multivariate.

Univariate outliers occur when a single variable is considered, for example, Sensor 1 Maximum Divergence. Firstly the data was converted to standard scores, which have a mean of zero and a standard deviation of one. This standardisation enabled the comparisons to be made easily across variables. Each element of the variable was then compared with the univariate distribution of the feature. If the element lay significantly outside the univariate distribution, standard scores of 3 to 4 [77], then the element was flagged as a possible outlier. The outliers were then analysed further to determine whether corrective action was to performed.

The univariate outlier tests revealed a large number of significant outliers in the data which can be seen in figure 4.5 for a typical response. Analysis of the outliers revealed that they were randomly distributed across the classes. Closer examination revealed that hard disk access caused interference to the data acquisition card. These multiple outliers were corrected so that they would not influence future results. For example, a voltage spike causing an inflated feature value could affect neural network scaling functions which use the maximum and minimum of a feature range order to re-scale for input to a neural network, usually 0 to +1 or -1 to +1.

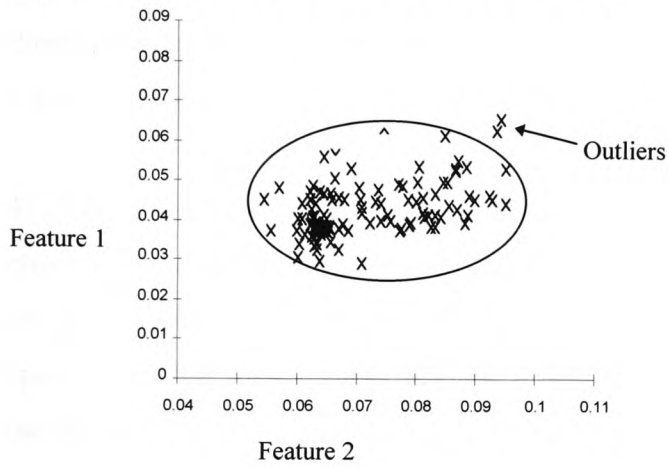
Figure 4.5 Cause of Univariate Outliers : Hard Disk Access



Pairs of variables were assessed for bivariate outliers by use of scatterplots. Elements which lay outside the range of the other observations were noted as

bivariate outliers. In order to simplify outlier detection an ellipse representing a confidence interval of the distribution, 90% in this case, was applied around the data point. Several extreme bivariate points were observed but were not removed from the population as they merely represented normal variations in the data. Figure 4.6 shows a scatterplot with two such outliers and their position relative to the 90% ellipse.

Figure 4.6 Scatterplot for Two Features Indicating Two Outliers



Since the data consisted of more than two variables, multivariate assessment of each observation was then performed. The multidimensional position of each observation was be measured relative to a common point. The Mahalanobis D squared method was used which measured the position of each observation from a multidimensional mean; a value of 0.001 was used as a threshold for outlier designation [78]. Several extreme observations were noted in the multivariate tests but again were not removed as they represented normal variations in the data.

4.2 Feature Analysis

The resultant data following the operations of data acquisition and pre-processing was an array of numbers, which represented the odour sample obtained from the sensor array. This large array of numbers may have in itself constituted the description of the sample. However, in this case there were too many numbers, many of which may not have helped to distinguish the current class from other classes. Feature Analysis served two main purposes: firstly it reduced the dimensionality of the problem and secondly it rendered the data more suitable for a decision process.

The general purpose of feature analysis is to determine the major components of classes which enable them to be distinguished from one another. Features which are irrelevant to the quality to be recognised need not be reflected in the pattern space. The data produced is also in a compressed form which eases further manipulation in order to make a decision. It must be noted however that the resulting features must retain sufficient information to enable patterns from differing classes to be distinguished. The features extracted must be represented in a reliable format; a set of features which changes significantly due to insignificant changes in the raw data would have limited the comparability of the features among samples.

The most significant features were determined in order to optimise any future pattern recognition decision. Merely combining all possible features obtained from the data series would not have necessarily increased the success of the system in distinguishing between classes [79]. Conversely, the grouping together of too many features might have decreased the pattern recognition efficiency.

The key to selecting features is, therefore, to determine the features which best distinguish one class from another. There were no equations to determine the best features from the data set thus each combination was evaluated separately to assess its feasibility for use in the final pattern recognition system [80]. The requirements of the feature selection process can be summarised as follows: the resultant set of

features must have low dimensionality, the features must have retained sufficient information, the distance in pattern space as a measure of similarity of physical patterns must be enhanced and the features of different samples must be able to be compared.

4.2.1 Feature Selection

The human odour sensing system incorporated an array of sixteen sensors from which sixteen features were derived for each sensor, as described in the section 4.1.3. This represented a substantial reduction in pattern space but was still too large and could have contained redundant or non useful information. The following strategy, to methodically assess the merits of each feature, was then undertaken.

The starting point for the feature selection algorithm was a large set of attributes for each individual sensor. Each sensor was assessed independently to ascertain whether dominant features were similar regardless of the sensor in question. The relative performance of each sensor in the array was then assessed after the best individual features had been chosen.

Following the measurement of the feature attributes, as described in section 4.1.3, the main selection algorithm was performed in order to produce a list of the best features according to a pre-defined selection criteria[90]. Three separate types of selection criteria were used in this analysis :

- Minimum error estimation
- Inter-class distance
- Probabilistic distance measures

4.2.1.1 Minimum Error Estimation

Minimum error estimation performed pattern recognition to each of the feature attributes. The data available was segregated so that a proportion of the data was used to construct a pattern classification rule and the remainder was used to test the discrimination capabilities of the rule. A score was therefore produced for each attribute which corresponded to the success of the pattern recognition scheme applied. The attributes which achieved higher recognition rates were deemed to be the most successful features. The pattern classification mechanism used was 10-Nearest Neighbour using 70% of the attribute data for developing the classification rule and the remaining 30% for testing the rule.

4.2.1.2 Inter-Class Distance Measures

The second selection criteria used was inter-class distance. This method used a pre-defined distance measure in order to calculate the distance from the feature attribute of one class to the feature attribute of another. The distance was averaged for measurements of a particular feature attribute between two specific classes. Therefore, if the inter class distances for an attribute were unreliable then the overall distance measure would reflect this unreliability. The distance measures used were [81] :

- Euclidean distance

Defined as :

$$\left\| F^i - F^j \right\|_{Euclidean} = \sqrt{\sum_{n=1}^{N} \left(F_n^i - F_n^j \right)^2} \tag{4.23}$$

where

N = the number of samples involved in the distance measure.

- Minowski distance

Defined as :

$$\left\|F^i - F^j\right\|_{Minowski} = \left[\sum_{n=1}^{N}\left|F_n^i - F_n^j\right|^s\right]^{\frac{1}{s}} \tag{4.24}$$

where

N = the number of samples involved in the distance measure.

- Chebychev distance

Defined as :

$$\left\|F^i - F^j\right\|_{Chebychev} = \max\left(n = 1, N\right)\left|F_n^i - F_n^j\right| \tag{4.25}$$

where

N = the number of samples involved in the distance measure.

An additional criteria of within-class distance was also included in this search strategy. The selection algorithm also graded attributes depending upon how the feature attributes for specific classes were distributed. If all of the attributes for a class covered a large range then the attribute score was reduced. The ideal feature would have provided a large inter-class distance to aid pattern classification whilst retaining a small within-class distribution to avoid overlapping with other classes.

4.2.1.3 Probabilistic Distance Measures

The final feature selection criteria used was probabilistic distance measures. This method assumed *a priori* statistical distributions of each of the feature attributes. These distributional assumptions were then used to more accurately calculate the distance between feature attributes of different classes in a similar manner to the inter-class distance method. The drawback with this technique was that if the *a priori* statistics were not readily applicable to the feature attributes in question, then the results would be inaccurate. The probabilistic measures used were[81]:

- Mahalanobis distance

Defined as :

$$\left\| F^{ci} - \mu \right\|_{Mahalanobis} = \sqrt{\left(F^{ci} - \mu\right)^T C^{-1} \left(F^{ci} - \mu\right)} \tag{4.26}$$

where

μ is the multivariate mean.

C is the covariance matrix.

- Matusita distance

Defined as :

$$D(x)_{Matusita} = \left\{ \int \left[\sqrt{p(x|c_1)} - \sqrt{p(x|c_2)} \right]^2 .dx \right\}^{\frac{1}{2}} \tag{4.27}$$

- Bhattacharyya distance

Defined as :

$$D(x)_{Bhattacharyya} = -\ln \int \left[p(x|c_1).p(x|c_2) \right]^{\frac{1}{2}} .dx \tag{4.28}$$

4.2.1.4 Feature Selection Results and Discussion

The three preceding sections presented separate methodologies for selecting the best features based upon distance metrics or classification rules. Each of the methods used provided certain characteristics which may have benefited a certain feature set, for example the probabilistic methods would have provided more favourable results for appropriately distributed features. The most successful features were determined by grading each feature according to the rank allocated for each feature assessment algorithm. The features accumulated a score according to the position of the feature after each individual assessment, zero for 'best feature' to fifteen for 'worst feature'.

The total accumulated feature scores for sensor 1 are shown in figure 4.7. The most successful five features were numbers 3, 2, 0, 5 and 7 which corresponded to

normalised maximum divergence, normalised standard deviation, normalised mean, normalised maximum absorption gradient and normalised maximum desorption gradient. The decrease in performance rating is 26% between the best feature and the fifth best feature which implies that the first five features all contribute discriminatory information. The worst three features were numbers 15, 10 and 13 corresponding to standardised absorption-desorption ratio, normalised absorption-desorption ratio and standardised maximum desorption gradient. The decrease in performance rating between the best and the worst feature is 156% which implies that the worst feature contains considerably less discriminatory information than the best feature.

The features which were calculated from standardised data, as defined in equation 4.5, performed considerably worse than the comparable features which were calculated from non-processed data, as defined in equation 4.3, but were then normalised across the entire array, as defined in equation 4.4. For example, the standardised maximum divergence was ranked at 8th, illustrating a 57% decrease in performance rating, compared to the normalised maximum divergence ranked 1st. Standardisation caused information loss as a result of the equalisation in sensor responses which occurred when the sensor responses were processed separately. The response inter-relations levels were, therefore, reduced using standardisation. Contrarily, the normalisation procedure was applied across the whole array of like features and resulted in the retention of sensor inter-relations but a simultaneous reduction in experimental errors in the odour concentrations. It can hence be concluded that the feature inter-relations between different sensors contain high levels of discriminatory information.

Figure 4.7 Feature Scores for Sensor 1

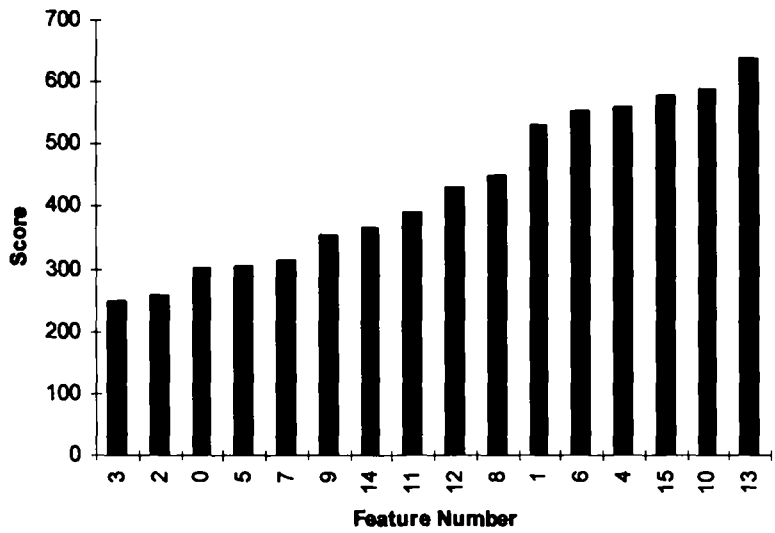


Figure 4.7 gives results for sensor one, but is indicative of findings from the other sensors when the five most successful features are considered. Since each sensor produced features of differing effectiveness it is advisable to assess sensor features individually to ensure that the features showing the highest discrimination are chosen. However, several sensors produced significantly different results which exhibited a closer scoring structure. These sensors were responding in a similar manner, as previously indicated in figure 4.4a, which provided very little discriminating power and consequently unusual feature selection results. The contribution of the actual sensors to the discriminating power of the system will be discussed in section 4.2.2.

Once the best features had been selected it was important to correlate each of the feature attribute pairs over the population in order to ensure redundancy was reduced to a minimum. If a high correlation was revealed between two attributes then they essentially contained the same discriminating properties, the lower priority feature was then discarded in such cases. Analysis of the best five features showed that the best three features (normalised maximum divergence, normalised standard deviation and normalised mean) were strongly correlated, greater than 0.96, and hence contained very similar discriminating properties. The features of normalised standard deviation and normalised mean were therefore eliminated to avoid redundancy of data. The remaining three features (normalised maximum

divergence, normalised adsorption gradient and normalised desorption gradient) also showed high correlations, greater than 0.70, but still contributed individual discriminatory information.

The set of sixteen feature attributes, per sensor, was therefore reduced to the following optimised set of three features :

- Normalised Maximum Divergence (F_3)
- Normalised Absorption Gradient (F_5)
- Normalised Desorption Gradient (F_7)

The effectiveness of the feature selection process was evaluated using the Matsuita Affinity as a performance measure[73], which is illustrated in figure 4.8. The Matsuita Affinity can be defined as :

$$\overline{S_M} = \frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} S_{M2}(f_i, f_j) \qquad (4.29)$$

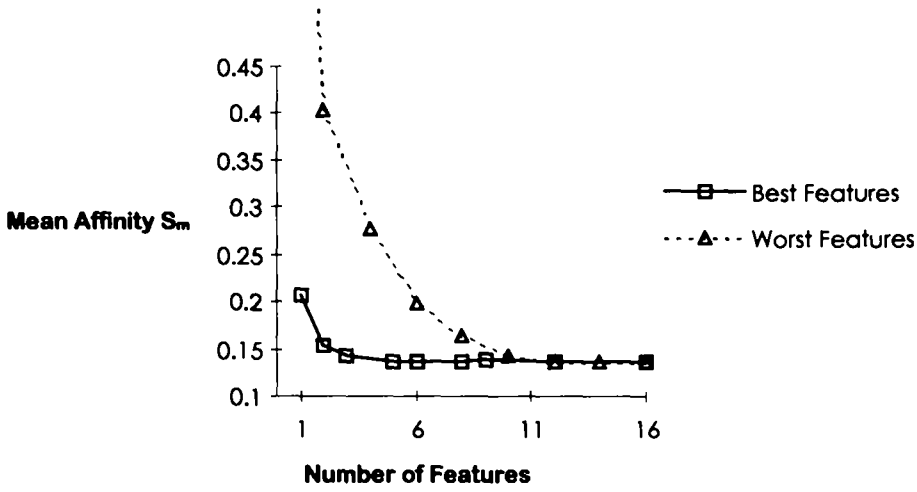
where

N is the number of features.

S_{M2} is the two class Matusita Affinity.

Figure 4.8 illustrates that the performance difference between the best selected features and the worst selected features, as shown in figure 4.7, can be easily seen. It can be seen from figure 4.8 that the discriminating power of the best feature set did not actually increase as inferior features were removed but merely remained at the same level until only the three chosen features remained. A reduction in discrimination was observed when normalised adsorption and normalised desorption features were removed. This reduction in discrimination implied that there was a significant contribution from the normalised absorption and normalised desorption features despite the high correlation with the best feature, normalised maximum divergence

Figure 4.8 Best and Worst Feature Sets (calculated for 2 classes) for Sensor 1, using Matusita Affinity, S_m, as the Feature Performance Measure



4.2.2 Sensor Assessment and Array Reduction

The discrimination capability of each of the sixteen sensors in the array was assessed using the previously selected optimised feature set of maximum divergence, absorption gradient and desorption gradient. A sensor was classified as unnecessary if it fell into one of the two following categories :

- Redundant
- Lacks Discrimination

A sensor which exhibits redundancy may still provide data which aids the discrimination capability of the system but which contains similar information to that of another sensor. Sensors were assessed for redundancy by analysing the correlations between pairs of sensors. If a pair of sensors exhibited a high correlation, greater than 0.96, then the sensor with the lower discrimination was discarded. This method does not usually increase the discrimination performance of the system but reduces the dimensionality of the data set which simplifies further pattern recognition techniques. This technique revealed a number of sensors which exhibited redundancy. This resulted in discarding a total of seven sensors from further analysis.

A sensor which exhibits a lack of discrimination is a sensor which does not contribute towards the effectiveness of the array in distinguishing between the different classes involved, people in this case. These sensors may also reduce the performance of the array by supplying conflicting class information caused by malfunction, noise or response to non-discriminatory odours. Analysis of the remaining sensors in the array revealed a total of four sensors which lacked sufficient discrimination information. Two of these sensors suffered complete failure during the course of the field trial. This failure caused a reduction in resistance of the sensor which shifted the baseline out of range of the data acquisition system. The failed sensors were irreversibly affected showing similar instability to the I-V characteristic illustrated in figure 4.2b. The cause of this transformation to instability is unknown but possible causes are; susceptibility to poisoning caused by overexposure to volatile odours, corruption of polymer chain by heating effects or poisoning, aging or polarisation due to d.c addressing technique.

The separation of the sensors in gas feature space [82] was analysed in order to verify and illustrate the sensor array reduction. This was performed by considering a single class, or person, and by determining the two dimensional discriminants for each sensor using multiple discriminant analysis, which will be discussed in further detail in section 4.2.3.2. This produced a cluster for each sensor on the two dimensional discriminant gas space[82]. Figure 4.9 shows this gas space for every sensor in the array used for the field trials.

Figure 4.9 re-enforces the sensor redundancy previously revealed by analysing correlations between sensor features. The discriminant plane clearly shows an area of very high density which represents the overlapping of information between various sensors of the array.

Figure 4.9 All Sensors Represented in Discriminant Gas Space

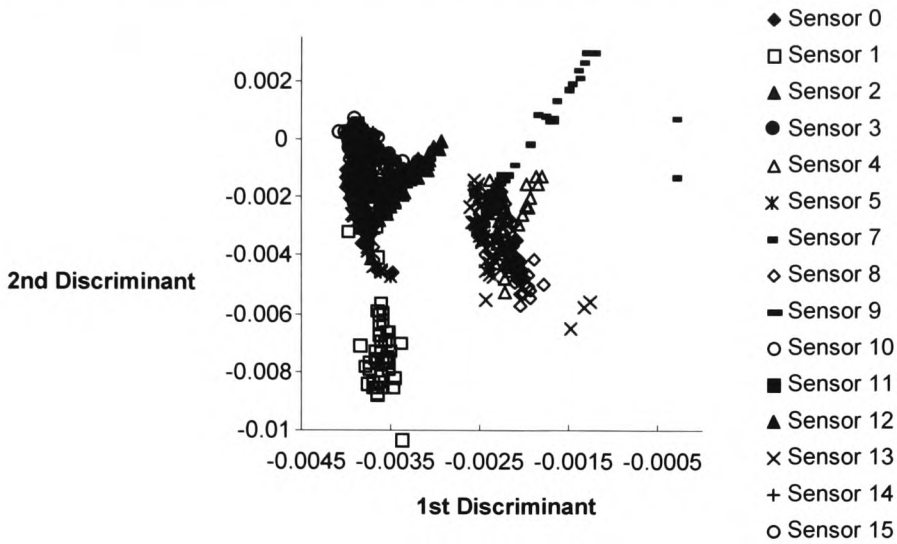
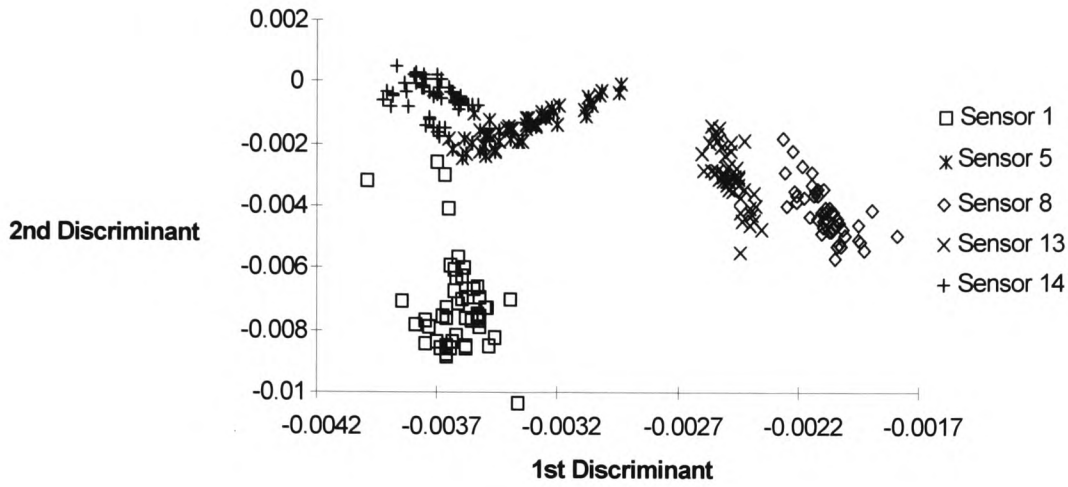


Figure 4.10 shows the two dimensional gas space for the same class, but only displays the optimised sensor array as determined previously. Figure 4.10 illustrates that the resultant array showed little redundancy between sensors with each sensor occupying a separate cluster on the discriminant plane.

However, the optimised array still showed a certain degree of interference between sensor responses, especially sensors 5 and 14, which implied the separate sensors were partially responding to similar groups of chemical odour compounds. However, these separations in discriminant gas space are given for a single class only and different classes showed varying degrees of separation between the sensors, depending upon individual odour profiles. This was illustrated by considering person 5 who demonstrated a higher degree of separation between sensors 5 and 14 but showed a reduction in separation between the remaining three sensors.

Figure 4.10 Best Five Sensors in Gas Space



The dimension of the sensor array was therefore reduced from 16 sensors to 5 by eliminating sensors showing redundancy or poor discrimination. The resultant array maintained the discrimination performance but greatly reduced the data needed for further pattern recognition algorithms. This decrease in data reduced the complexity of the system whilst simultaneously decreasing learning times. The effectiveness of the optimised array will be investigated further in chapter 5.

4.2.3 Feature Extraction for Exploratory Data Analysis

Feature extraction differs from feature selection in that it actually produces transformations of the input data in order to maximally separate pattern classes and to reduce the dimensionality of the feature space. This differs from feature selection which attempts to provide an optimal feature set from a given number of feature attributes, no transformations are performed in those procedures. There are many different feature extraction algorithms available but they can broadly be categorised into linear and non-linear methods. Both Linear and non-linear techniques were used in order to gain a more thorough knowledge of the data structure and discriminating capabilities of the human odour sensing system developed.

Feature extraction was used for several purposes :

- Exploratory data analysis to assess the discrimination capabilities of the system
- Assessment of the extent of linearity / non-linearity of the system

The input to the feature extraction algorithms was the optimised feature set, as detailed in section 4.2.1. The success of these feature extraction techniques were therefore dependent upon the suitability of the input features. The software and methods used to perform the feature extraction processes prohibited the use of high dimensional input vectors. The output from the algorithms was a two dimensional feature vector which attempted to represent the 15 dimensional input space, 5 sensors with three features per sensor, by applying a transformation to the input space. A two dimensional vector was chosen in order to provide a visual representation of the class separability by utilising scatterplots.

The feature extraction techniques determine class relationships and patterns in either a supervised or unsupervised manner. Unsupervised means that the input data is supplied unlabelled so that the particular algorithm operates as a clustering algorithm which produces a 2 dimensional class vector. Supervised extraction methods use the class membership data as well as the unseen data in order to determine the resulting pattern separation space. However, the algorithms do not make any predictions on the number of classes present or any explicit class predictions.

These methods were suited to the exploratory nature of this particular study in which class clustering and other class related observations were assessed. The data format utilised for this exploratory analysis was the optimised feature set derived from section 4.2.1.4 which contained a total of fifteen features, 3 features per sensor, for every odour sample. The features were subjected to a linear normalisation(0 to 1) which is similar to a neural network input transformation. The majority of the data considered is presented for two classes only, which simplifies the visual appearance of the data in two dimensions, although more classes are considered when appropriate.

Four different techniques were chosen, each of which was intended to provide a different insight into the characteristics of the data. The reason for the choice of algorithm and the insights to be gained are described at the beginning of each feature extraction section.

4.2.3.1 Principal Components Analysis (PCA)

Principal Components Analysis is a linear technique which is unsupervised in nature, meaning the class membership information is not utilised when calculating the components. This technique was chosen to determine whether the two classes were linearly separable, without the use of the class membership data. If the classes prove to be linearly seperable then linear pattern recognition methods could be utilised which would greatly simplify future procedures. The unsupervised nature of this technique ensured that results were not influenced by prior knowledge of class membership and hence showed the true structure within the data itself.

PCA allows a large set of variables to be reduced into a substantially smaller set of uncorrelated variables. This reduced set represents most of the information in the original data set and is much easier to understand and use in further analysis. The idea was originally conceived by Pearson(1901) and independently developed by Hotelling(1933)[83]. The uncorrelated variables, derived from the feature set of each sample, maximise the variance accounted for in the original variables and are called the principal components. If two principal components account for the majority of the variance in the original fifteen features then the dimensionality has been reduced from fifteen correlated dimensions to two uncorrelated dimensions.

If the principal components are calculated from the covariance matrix then the first principal component is dominated by the variable with the largest variance since this direction maximises the variation. Therefore, using the covariance matrix would be an advantage if a large sensor response implied a strong correlation with the individual being sampled. This assumption, however, is dangerous for several reasons; if the response of one particular sensor is comparatively very large then the response of the remaining sensors will be masked. This danger could also reduce the discrimination between people if a particular sensor gives a very large

response irrespective of the person being sampled. Hence each person, in this case, would produce a similar first principal component. The problem of greatly differing variances can be eliminated by using a correlation matrix to calculate the principal components. Correlation provides a measure of how closely two variables move together, irrespective of absolute magnitudes involved. It can be shown [84] that this definition of principal components leads to the matrix equation :

$$\boldsymbol{Ra} = \boldsymbol{\lambda a} \tag{4.30}$$

where

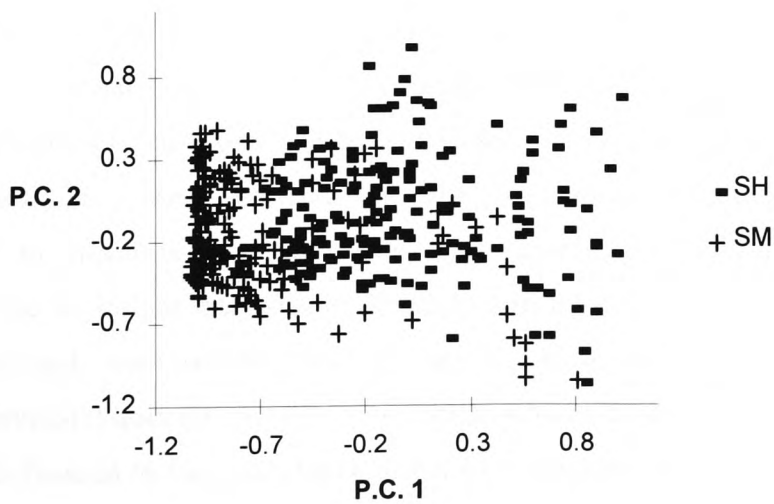
$\boldsymbol{\lambda}$ is the eigenvalue of the correlation matrix $\boldsymbol{R}$.

$\boldsymbol{a}$ is its associated eigenvector.

This matrix equation can be solved for $\boldsymbol{\lambda}$ and $\boldsymbol{a}$, the basic statistics of principal components. If the correlation matrix, $\boldsymbol{R}$, is non-singular (there are no exact linear dependencies among variables) then there are $\boldsymbol{p}$ eigenvalues, $\boldsymbol{\lambda i}$, and $\boldsymbol{p}$ associated eigen vectors, $\boldsymbol{ai}$, that satisfy the equation.

Figure 4.11 shows the first 2 principal components for two classes, or people, from samples taken over the entire sampling period of 6 weeks.

Figure 4.11 First 2 Principal Components for SH and SM classes



In figure 4.11 the separate clusters representing the two classes can be easily observed but the data points merge into one another implying a certain degree of non-linearity between the classes. This 2-dimensional representation illustrates that there was significant confusion between the two classes, persons SH and SM.

The PCA method performs a linear orthogonal projection where the two dimensions of the graph represent the orthogonal vectors which show the largest and second largest variances. The vector showing the highest variance is termed the first principal component and figure 4.11 clearly shows that it, labelled P.C.1, describes the majority of the variance of the data as the cluster separability is mainly attributable to the vertical separation of the points.

The PCA method shows that the data supplied for user SM demonstrated a lower degree of within-class spread than the user SH. This can be seen from figure 4.11 by the smaller and more clearly defined cluster exhibited by user SM. However, although PCA attempts to maximally preserve the variance of the data it does not maximally preserve the inter-pattern distances. The Sammon map used in section 4.2.3.3 provides more reliable cluster information.

4.2.3.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a linear technique which is supervised in nature, meaning it utilises class membership data in order to optimise the discriminant calculations. This technique was chosen to determine whether the two classes were linearly separable when the class membership information was also used to maximise class separation. This supervised technique was chosen to provide an insight into whether linear pattern recognition techniques, which are supervised, were suitable for this data set. However, since this technique is supervised it does not provide a true insight into the data structure since the results are influenced by the prior class membership knowledge to improve separability.

LDA implements a linear projection but unlike PCA it is not necessarily orthogonal. LDA attempts to find a plane in the high dimensional feature space and then alters the plane to maximise class separation[81]. However, this method does use *a priori* information regarding the class distributions and hence is regarded as a supervised technique. This prior knowledge aids in the ability of the method to separate the classes.

Figure 4.12 Linear Discriminant Analysis for Classes SM and SH

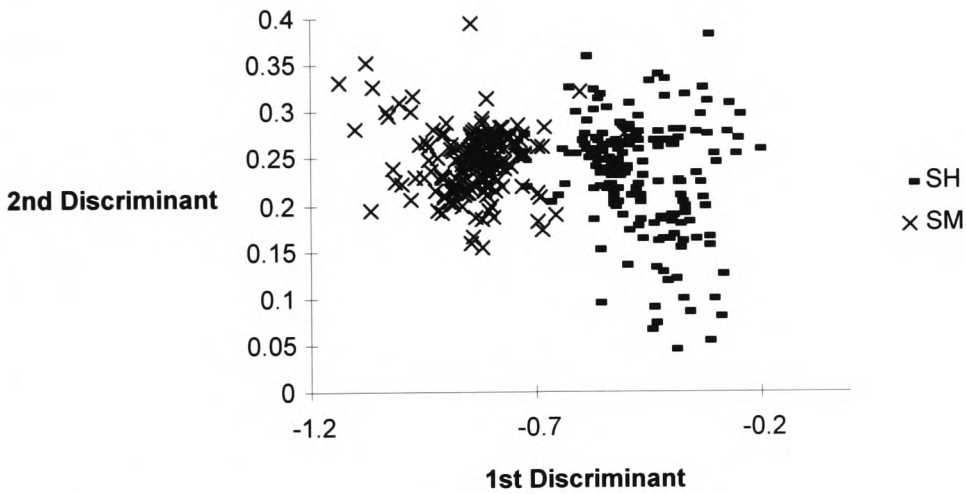


Figure 4.12 shows an almost linearly separable pair of clusters representing the two classes under consideration. Even with the use of prior statistical knowledge the clusters are still not completely linearly separable as some overlapping can be observed. The majority of the discriminatory information again is contained in only one dimension, the 1st discriminant in figure 4.12.

The cluster shapes again imply a slight increase in spread for the user SH but LDA does not retain inter-pattern distances because class separability is the primary goal of the algorithm. The Sammon method in the following section provides a map which retains the inter-pattern distances and so gives a more reliable insight into cluster shapes.

4.2.3.3 Sammon Map (SAM)

Sammon mapping is a non-linear technique of unsupervised nature, meaning the class membership data is not used whilst constructing the two dimensional map. This technique was chosen to determine the whether the two classes were separated non-linearly without the use of class membership data. The results produced by this technique could then be compared to the un-supervised linear method of PCA to determine if class separability was improved using a non-linear algorithm. This method also provides a more reliable estimate of the actual within class spread and inter-class distance since this method attempts to maximally retain the structure of the original data set, this method is hence described as being a topographic map.

The Sammon map operates in a similar manner to the PCA method but uses its non-linearity to preserve the structure of the data when transformation occurs[85].

Figure 4.13 Two Dimensional Sammon Map for Classes SH and SM

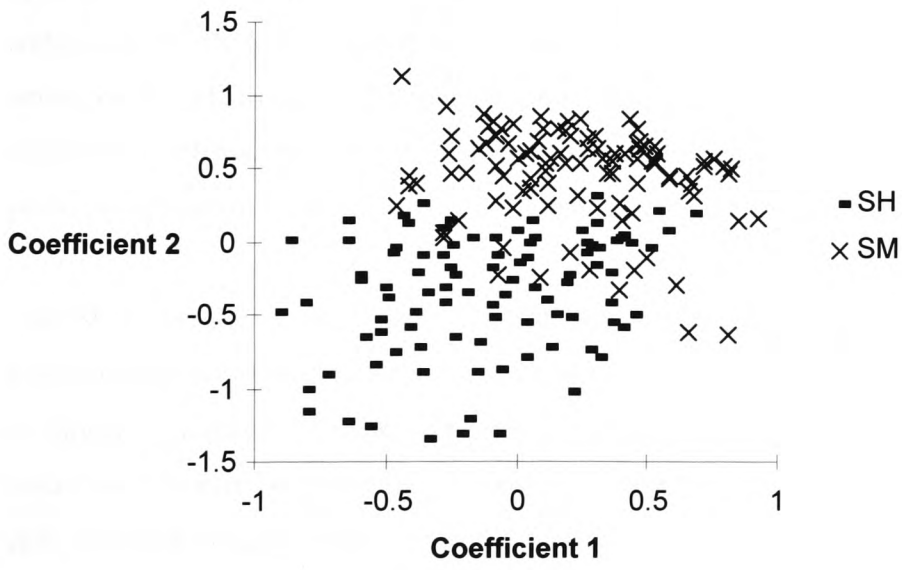


Figure 4.13 again clearly shows the two separate clusters but still exhibits considerable overlap between the two classes which implies some non-linearity in the data. Similarity between the PCA and the Sammon map can be visualised except for the shifting of the plane of separation.

Since the Sammon map attempts to maximally preserve the inter-pattern distances of the data the true cluster shapes of the two classes can be seen from the graph. The two clusters are both similar in size and shape but the cluster of user SH shows a greater degree of spread than that of user SM. An increase in cluster spread may be attributed to various causes, for example more unreliable sample position or less stable odour profile.

4.2.3.4 Kohonen Self Organising Map (SOM)

The Kohonen Self Organising Map (SOM) produces a non-linear transformation of the measurement space, unlike PCA which produces a linear transformation. This method also operates in an unsupervised manner, meaning that the class membership information is not utilised whilst feature extraction occurs. This technique was chosen for the same criteria as the SAM technique; it is un-supervised, non-linear and produces a topographic map. However, this technique operates in a substantially different manner to the previous three statistical techniques of PCA, LDA and SAM, because it is a neural network. Neural networks are able to offer enhanced pattern recognition performance in linearly in-separable problems. The SOM was therefore chosen to provide an insight into the possible application of neural networks for pattern recognition purposes.

The SOM is able to discover statistical regularities in its input space and automatically develops different modes of behaviour to represent different classes of inputs. A Kohonen feature map only possesses one layer of neurons and all inputs are connected to all nodes. The learning algorithm organises the nodes in the grid into local neighbourhoods which act as feature classifiers on the input data. The topographic map is autonomously organised by a cyclic process of comparing input patterns to vectors 'stored' at each node. No training response is specified for any training input. Where inputs match the node vectors, that area of map is selectively optimised to represent an average of the training data for that class [86].

The Kohonen layer determines the winning Process Element, PE, by computing a distance between each PE's weights and the input value. This distance is usually the Euclidean distance between the two vectors and is given by [86]:

$$D_i = \sqrt{(x_1 - w_{i1})^2 + (x_2 - w_{i2})^2 + ...+(x_m - w_{im})^2}$$

The input layer has m values represented as :

$$X = (x_1, x_2, ..., x_m)$$

and hence each Kohonen PE will also have m weight values and can be denoted by

$$W_i = (w_{i1}, w_{i2}, ..., w_{im})$$

A conscience mechanism adjusts the distances to encourage PEs that are not winning above an average frequency and to discourage PEs that are winning above an average frequency. This helps develop a uniform data representation in the SOM layer.

The adjusted distance is simply the distance minus a bias, which is computed using the following :

$$B_i = \gamma(N.F_i - 1)$$

where

F_i = frequency with which the PE has historically won.

N = number of PEs in the SOM layer.

γ = Learning rate of network.

Figure 4.15 Kohonen SOM for Classes SM and SH

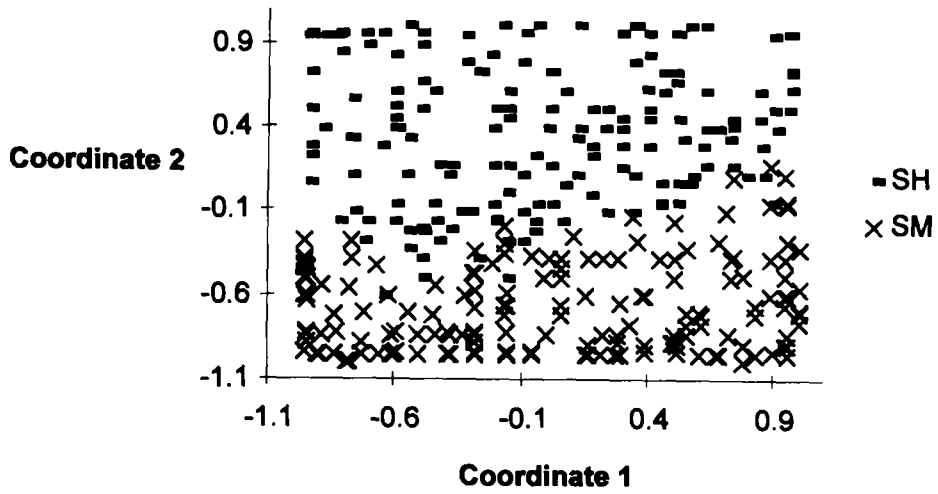


Figure 4.15 shows a two dimensional Kohonen SOM, using a 10 x 10 neuron architecture, for the same two classes as used by the previous feature extraction methods. The map illustrates two clear clusters representing the people SM and SH. The map exhibits a small degree of overlap and a distinctly non-linear boundary between the classes is present. It must be noted that the patterns are uniformly distributed in the two dimensional SOM space when compared to the other methods utilised. This is a distinctive attribute of the SOM and could render cluster analysis almost impossible if prior class attribution is not known, however the class attributions are known in this case so data points can be easily attributed to the correct class.

The SOM is similar to the Sammon map as both are termed topological maps. This means the maps attempt to retain the inter-pattern distances when the features are transformed. The SOM therefore can be used to analyse the shape of the clusters and their corresponding spread. The SOM shows that the data points for SM are marginally more tightly spaced than those corresponding to SH, this result is in accordance with the Sammon map, presented in section 4.2.3.3.

4.2.3.5 Human Response Variation

The previous figures in this feature extraction section have shown that there was significant interference between sensor responses for the two classes used. These results were produced using odour response data taken throughout the entire sampling period. Figure 4.16 illustrates a SOM produced for the same two classes but this time using only the sampled data for one day only.

Figure 4.16 Two Class (SH and SM) Separation for One Day Only

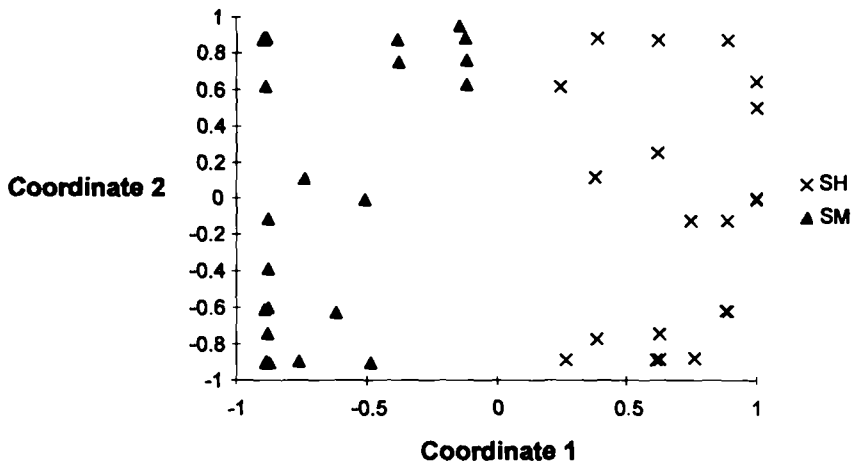
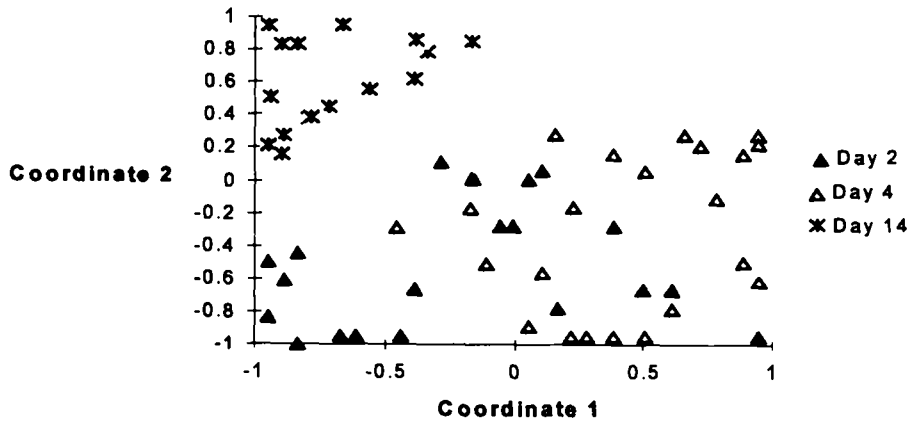


Figure 4.16 clearly shows a large degree of separation between the two classes when only one day is considered. When compared to figure 4.15, which shows long term data for the two classes, it can be seen that as time increases the samples drift in feature space with resultant interference between classes. The results shown in figure 4.16 can be replicated for almost any single day which would result in almost perfect discrimination if each day were treat as a new starting point.

Figure 4.17 illustrates how the response, in transformed feature space, for a single class drifts with respect to time. The SOM plot shows that within the time period of just two days the responses had only marginally moved in feature space, which can be seen from the two overlapping clusters for day 2 and day 4.

Figure 4.17 Response Variation with Respect to Time for Class SH (using SOM)



The final cluster shown in figure 4.17 represents the cluster, for class SH, from day 14. This cluster can be seen to be completely separated from the two earlier days and is in fact linearly separable. This separation shows that as time increases the sample data for a specific class drifted significantly from its starting point and the separation increased as time increased. Therefore, if all classes were considered then the various class clusters increasingly overlap with each other as time elapses.

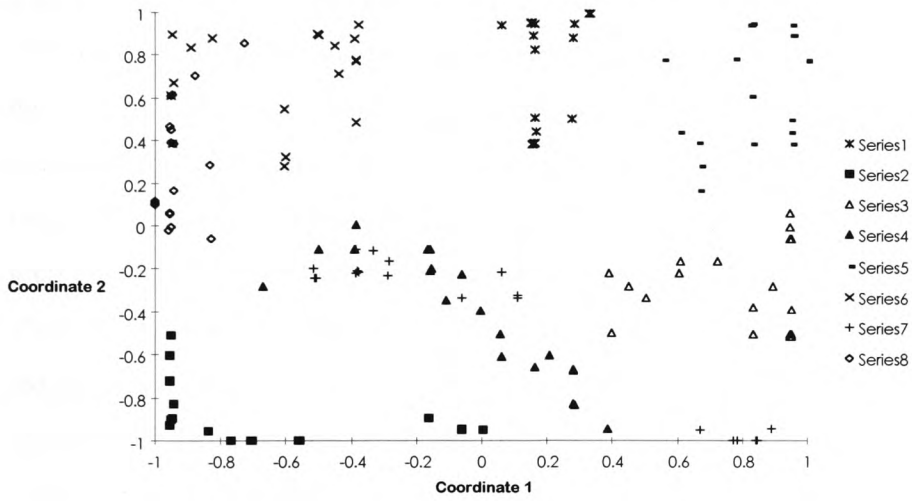
All classes exhibited shifts in class clusters as a function of time, some showed greater variations than others. However, not all classes behaved in the same manner as the class visualised in figure 4.17. Other classes exhibited a more random shift pattern with clusters moving around a defined area of feature space.

These results demonstrated that the odour patterns measured by the electronic nose showed significant drifts with respect to time. Attempts were made to correlate these drifts with external influencing factors but no direct correlation was found. The influence of temperature and humidity was discussed in section 4.1.2. The drifts were in fact relatively individual in nature which implies that the variations were caused by human variations or changes in sensor response mechanisms.

All of the results in this section have been concerned with two classes only in order to aid visual analysis of the two dimensional representations. As the number of classes increases the transformed pattern space increases in complexity as class clusters interrelated due to various factors influencing sensor response variations. The SOM plot in figure 4.18 illustrates the class separations for the eight people

who participated in the field trial. The results shown in figure 4.18 are produced for a single day only in order to simplify the map and also to reduce the cluster spread for ease of visual analysis. As previously discussed, the class clusters were reasonably separated, in the two dimensional space, when only a single day is considered. Figure 4.18 clearly demonstrates how the interference between a minority of classes, for example classes 4 and 7, caused the majority of the misclassification.

Figure 4.18 SOM for All 8 Classes Over a Single Day Duration



4.3 Discussion

Section 4.1 of this chapter began by considering the representation of the un-processed digitised sample data, collected during the six week field trial. The aim was to reduce the dimensionality and redundancy of the data set but retain the discriminatory information. This standard optimised data set was intended for use in subsequent techniques of feature extraction, see section 4.2, and pattern recognition, see chapter 5. The various techniques employed could then be readily compared since the input data set was always consistent. A set of 16 features were presented and were calculated for every sensor and every sample. These features effectively reduced the sample data size from 370 data points for each of the 16 sensors to 16 one dimensional feature attributes for each of the 16 sensors. The discrimination ability of the 16 features was then assessed which revealed that just

3 features per sensor represented the majority of the discriminatory ability of the data. The addition of any of the remaining 13 features would have merely duplicated discriminatory information or provided redundant information which could have reduced the performance of subsequent pattern recognition techniques.

This optimised feature set, 3 feature attributes per sensor per sample, was then used to assess the individual discriminatory contribution of each sensor in the 16 dimensional sensor array. Analysis showed that two sensors suffered complete failure and hence did not contribute any discriminatory information. A further two sensors responded well to human odour compounds but did not contribute any information which would have helped discriminate between the people involved in the field trial, this was due to similar responses obtained for all people involved in the field trial. This common response could have been caused by the sensor responding to common human odour compounds instead of unique compounds. A total of twelve sensors were selected which contained discriminatory information. This sensor subset was then reduced further to a subset of five sensors to eliminate redundancy, since seven of the sensors contained similar discriminatory information. Each odour sample was then represented as an array of fifteen one dimensional features, three features calculated for each of the five selected sensors. This array of fifteen features was then used for all subsequent feature extraction techniques. This optimised feature/sensor subset will also be used for all of the pattern recognition techniques applied in chapter 5.

Section 4.2 investigated several feature extraction algorithms in an attempt to determine any characteristics within the data. The fifteen dimensional optimised feature/sensor set was used as the input to all of the algorithms. The data was restricted to two classes, people, to enable easy interpretation in a two dimensional plot of the extracted features. The data for the entire field trial period, six weeks, was initially used in the analysis.

Four algorithms were chosen, each intended to illuminate separate characteristics of the data. Principal Components Analysis (PCA) produced poor results which showed a large degree of overlapping between the two classes. It was deduced that

the data contained non-linear relationships which PCA was unable to separate because of its purely linear behaviour. The technique of Sammon mapping (SAM) began with a PCA phase but then progressed non-linearly to maximise the class separation. The SAM produced superior results when compared with PCA, which again implied that the data set contained non-linearities. However, despite the SAM's non-linear behaviour a large degree of overlap between the two classes still existed. The Kohonen Self Organising Map (SOM) produced far superior separability when compared to the techniques of PCA and SAM. The SOM was a non-linear technique but differed greatly to the SAM since it possessed a Neural Network architecture. The two classes were clearly visible with low overlap, a non-linear boundary, however, still existed between the two classes. It was deduced from these results that a non-linear pattern recognition method would provide optimum results. The SOM results also suggested that neural computing methods could enhance pattern recognition performance.

The technique of Linear Discriminant Analysis (LDA) was also applied to the optimised data set. This technique differed from PCA, SAM and SOM because it actually used a priori class information to maximally separate the data whereas the previous three techniques merely searched for patterns within the data without the use of a priori information. Despite the purely linear nature of this technique it produced two easily definable clusters for the two classes with only minimal overlap between the two classes. Although the success of this technique was attributed to its supervised nature it was deduced that linear pattern recognition techniques may also produce similar results since they are also supervised in nature.

It can be concluded that neural computing pattern recognition techniques are predicted to offer the highest discrimination between people as a result of the superior separability produced by the SOM, even when un-supervised in nature. Linear statistical techniques are predicted to offer lower discriminatory ability because of the high degrees of non-linearity in the data set. However, linear pattern recognition techniques should be included since LDA produced a high degree of separability and also as a means of measuring the performance of the neural

network techniques. The use of statistical pattern recognition techniques may also provide more insight into the structure of the data set.

Further analysis of the SOM data revealed that the two classes were more easily separated on shorter timescales. For example, the two classes were easily separated when a two day period was considered but classes became increasingly overlapped as time increased. This observation was confirmed for all people in the field trial. When all eight people were considered on a single day approximately 80% of the classes were separated but separability reduced when data from a larger timescale was considered. This time variance is a potential problem when considering pattern recognition, but the high separability when considering short timescales may be exploited by introducing an algorithm in the pattern recognition process, and will be investigated further in section 5.3.1.3.

5. Pattern Recognition

A number of pattern recognition experiments were performed on the optimised feature set in order to determine the effectiveness of the system at identifying or verifying a human based upon the odour profile supplied to the electronic olfactory device. These experiments were divided into two sections: Statistical Methods and Neural Computing. Both statistical and neural computing techniques were used as a result of the insights gained into the data structure following the feature extraction algorithms performed in section 4.2. Different techniques were performed in each of these areas in order to gain an insight into the data structure.

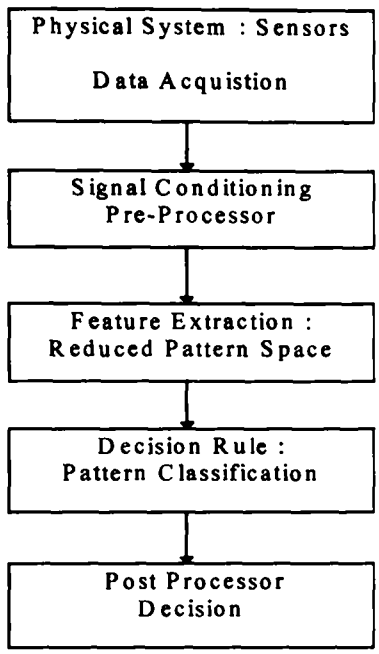
5.1 Introduction

A pattern recognition process involves two distinct parts: the formation of a decision rule and using the rule to classify a sample. The decision rule is developed using the labelled patterns in the learning phase. The resulting decision rule must possess various capabilities in order to be of true use in a pattern recognition system. The decision rule must accurately represent the data from which it is developed but pure representation of the learning data does not constitute a good pattern recognition system. A system which only correctly classifies exact copies of training data but rejects patterns with natural variations is a poor example of a such a system.

A good pattern recognition system, therefore, generalises the data used for learning; a decision rule is developed which is flexible enough to absorb natural variations in the structure of different patterns. For example, a bad pattern recognition system designed to recognise the letter 's' only recognises the letter 's' if the font of the learning data is presented. A pattern recognition problem thus begins with class definitions and labelled samples of those classes in some workable representation. The problem is solved when a decision rule is derived which assigns a unique label to new patterns [80].

The pattern recognition system can be subdivided into several distinct operations as illustrated in figure 5.1.

Figure 5.1. A Pattern Recognition System



The physical system, as shown in figure 5.1, represents the method of obtaining the patterns which are in need of classification. In the case of odour sensing, the physical system was represented by a finite array of conducting polymer gas sensors. The mechanism of data acquisition converts the physical dimension into a form which can be used in the further operations of the pattern recognition system. This acquired data is contained in measurement space.

The data in measurement space then undergoes signal conditioning and pre-processing which attempts to eliminate the effects of the data acquisition system, masking the true trends of raw data such as noise and the effects of amplification. The subdivision could involve many operations, or perhaps none, depending upon the quality of data received from the data acquisition system. Section 3.8 described this area in greater detail.

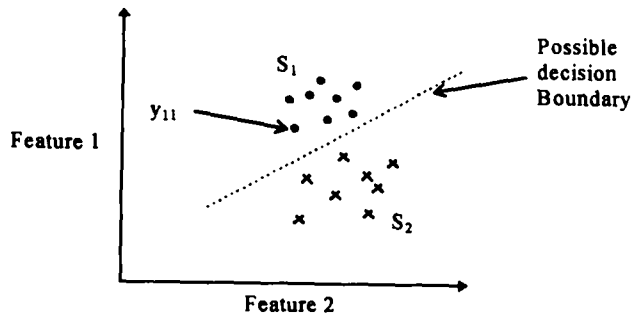
Following signal conditioning and pre-processing, the data then undergoes feature calculation. The overall aim of calculating features is to reduce and refine the data set to an acceptable size so that future processing is simplified or becomes a possibility. The actual features used in this analysis were selected using the methods presented in section 4.2.1.

Finally, a decision rule must be developed using a set of labelled samples, the training data, to enable the system to classify a point in the pattern space corresponding to an unlabelled sample. Once the decision rule is developed, it can be used to classify new patterns into a specific class. The output from the decision rule can be interpreted by further stages, using a post processor to convert the class information into an appropriate form. This information is often converted into a probability of class membership for a new instance of a pattern; the pattern recognition system can then act upon this probability.

A pattern can hence be described by a finite number of variables called features represented by $x_1, x_2, \ldots\ldots, x_n$. A specific pattern is a point $\mathbf{x} = (x_1, x_2, \ldots\ldots, x_n)$ in an n dimensional pattern space X which occupies the area in which all of the patterns can occur. There are a finite number of pattern classes $C_1, C_2, \ldots\ldots, C_i$ which are used to classify all of the points of the pattern space. The occupants of these classes are unknown except for a finite number of labelled samples, the training set, which are used to derive the decision rule associated with the pattern classification problem.

Pattern classification can be easily visualised if only two features and two classes are considered since the patterns can then be plotted on a graph and can be easily interpreted. Figure 5.2 shows a graph of two classes illustrating linear separability using the decision boundary shown.

Figure 5.2 A 2 Class 2 Feature Pattern Classification Problem



Pattern recognition is not, therefore, a single process but consists of many interlinked and interdependent processes. The following sections present a number of different pattern classification methods and the associated performance results based upon the odour test set obtained in the field trial. The methods differ in both complexity and also the manner in which the feature space is partitioned into the regions for each individual class. The majority of pattern classification techniques applied to olfactory arrays utilise neural computing techniques whilst standard statistical techniques are not usually considered. The emphasis on the following attempts at classification is to assess the merits of both statistical and neural computing techniques in order to determine the most suitable form of classification system for this particular application. The various techniques will therefore be compared directly at the end of this chapter.

The classification techniques are evaluated by use of several standard measures, in addition to performance measures specific to a classifier. The first two methods, FAR and FRR, are commonly used to measure the performance of a biometric *verification* device, as previously described in section 1.2. The methods of Recognition rate and Absolute Recognition Rate are applicable to a biometric *recognition* device. There are fundamental differences between biometric verification and recognition. Verification only considers the probability of class membership for the current user requiring authentication. The user supplies a template number by use of a suitable interface, for example a keypad or tag. However, recognition does not require any prior identification from the user apart from the biometric information. The system then provides the most suitable match

combined with the probability of class membership to the winning class. These measures can be defined as follows:

- False Acceptance Rate (FAR) as defined in equation 1.11. This is calculated by determining the number of class probabilities which reach the desired threshold rate, excluding the probabilities of the currently selected class. This then indicates the likelihood of the classifier to falsely verify a user, who is enrolled on the system but as a different template.
- False Rejection Rate (FRR) as defined in equation 1.12. This is calculated by determining the number of class membership probabilities, for a given output class, which reach the required threshold criteria. This measure then indicates the capability of the classifier to correctly recognise a legitimate user of the system.
- Recognition Rate (RR) shows the relationship between the percentage of correctly classified samples to the threshold level applied to the resulting class membership probabilities. This rate can be expressed mathematically as :

$$RR = \left(\sum_{k=0}^{c-1} f(P_k, m_k, n_k, \tau) \right) \cdot \frac{100}{N} \tag{5.1}$$

where

N = number of sample instances in the test set.

x = output probability threshold.

m_k = winning class for sample k.

n_k = actual class represented in sample k.

P_k = probability of winning class.

τ = threshold.

f is the decision function defined by :

$$(P_k, m_k, n_k, \tau) = \begin{cases} 1 \text{ if } m_k = n_k \text{ and } P_k \geq \tau \\ 0 \text{ if } m_k \neq n_k \text{ or } P_k < \tau \end{cases}$$

- Absolute Recognition Rate (ARR) is the percentage of correctly classified samples irrespective of threshold, hence the recognition rate at a threshold of zero. This rate is also called the hit rate of the system.

The FAR and FRR curves will be used to evaluate the performance of each classifier locally in the section relating to each of the pattern recognition methods, whilst the RR curves and the ARR will be used to compare the performance of the various methods at the end of the chapter, in section 5.4. The measures of FAR and FRR are considered locally because specific pattern recognition techniques can be analysed in detail using these curves. Comparison of all performance measures for every pattern recognition method would be very confusing, hence the RR curve is used for comparison since a single curve is only required for each method which reduces complexity and allows the methods to be compared clearly. The RR curve also directly relates to the ARR, which is a single percentage performance measure, and can be used to easily compare the performance of all methods clearly and simply, as shown in figure 5.20.

The data used to induce statistical classifiers and to train neural networks was identical irrespective of the technique concerned. The data was partitioned into 50% training, 10% validation and 40% test data which was split chronologically so that recognition on a time varying basis could be assessed. The validation data, including data from all classes, was reserved for verification in the training phase so that training could be monitored continuously, where necessary. The test set could not be used for this verification purpose because the resultant classifiers would be inadvertently biased towards the recognition of the actual test set samples. This data set shall be referred to as the 50/10/40 data set.

The standard data set was also partitioned in a different way, for the case of the back propagation neural computing technique, in order to illustrate the performance of a 'real life' biometric access control system which would only utilise training data from an initial enrollment period, for example the data supplied by the user on the first day of use. This data set was partitioned into 10% training, 5% validation and 85% test data which was again split chronologically so

that recognition on a time varying basis could be assessed. This partitioning of the data was a departure from the usual methodology of utilising 60% for training and 40% for test, but was chosen to provide recognition rates similar to those produced by a commercial biometric system. This data set shall be referred to as the 10/5/85 data set.

5.2 Statistical Techniques

As previously discussed, the aim of pattern recognition is to assign a new set of input data to one of a finite number of classes, people in this case. A new input pattern $\mathbf{x}$ must be assigned to one of c classes, C_k where $k = 1,...,c$. A benefit of statistical pattern recognition is that a theoretical best performance can be defined which corresponds to the lowest probability of misclassifying a new input pattern. The probability of misclassifying a new input pattern is almost always non-zero unless the classification problem is extremely simple [87].

The probabilities of a new input pattern $\mathbf{x}$ belonging to each of the c classes is given by $P(C_k \mid \mathbf{x})$. These are called the posterior probabilities since they are constructed following the assessment of the new input vector $\mathbf{x}$. If a large number of input vector observations are considered for a specific class C_k then the class conditional distribution can be written as $p(\mathbf{x} \mid C_k)$. The class conditional distributions correspond to probability density functions since the input vector $\mathbf{x}$ is continuous. The distribution of all of the input vectors irrespective of class is given by $p(\mathbf{x})$ and is called the unconditional distribution. The final distributions to be considered are the occurrences of the classes irrespective of the input vectors which can be expressed as $P(C_k)$. These are called the prior probabilities since they correspond to probabilities of class membership before the input vectors are observed.

These various measured probabilities can be combined to form Bayes Rule which is given by[87] :

$$P(C_k \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_k)\, P(C_k)}{p(\mathbf{x})} \tag{5.2}$$

where

$$p(x) = \sum_{k=0}^{c} p(x|C_k) P(C_k) \tag{5.3}$$

which ensures the posterior probabilities sum to one, this denominator is the same irrespective of class.

Hence, in order to minimise the probability of misclassification, the new input vector should be allocated to the class for which the posterior probability is highest. Since the probability p(x) is independent of the class, this class assignment protocol can be expressed as :

$$p(\mathbf{x} \mid C_k)P(C_k) > p(\mathbf{x} \mid C_j)P(C_j) \text{ for all } j \neq k \tag{5.4}$$

Pattern classifiers are rules, applied to new patterns, which assign that pattern to one of the c classes. For the purpose of the classifier the feature space can be divided into c decision regions, $R_1,...,R_c$. Theoretically, these regions may be divided into several separate regions to account for multimodal distributions but may be limited depending upon the classifier in question. If a new pattern falls in region R_k then the pattern is assigned to class k. A misclassification will occur if a pattern which, for a two class case, belongs to C_1 is assigned to C_2 or vice versa. The probability of either of these misclassification errors can be expressed by [88]:

$$P(error) = P(\mathbf{x} \in R_2, C_1) + P(\mathbf{x} \in R_1, C_2) \tag{5.5}$$

where $P(\mathbf{x} \in R_2, C_1)$ is the joint probability of $\mathbf{x}$ being assigned to region 1, and hence class C_1, whilst actually belonging to class 2. If a single feature, two class

problem is considered then the decision boundary which minimises the probability of misclassification is the crossing point of the two joint distributions.

Therefore, the pattern classification process involves estimating the underlying probabilities required to induce pattern classifiers. If the distributions are inaccurate then it follows that the induced classifier will also prove to be inaccurate at correctly assigning new patterns. In this work various statistical techniques were utilised in order to assess their discriminatory abilities for the field trial data.

Statistical classifiers can be broadly divided into the following two main areas :

- Parametric
- Non-Parametric

Parametric methods rely upon the input class conditional distributions conforming to a predefined statistical distribution, usually multinormal distribution. The parameters of this distribution are then calculated from the training exemplars. This method will obviously produce inferior results if the real distributions are unlike those of the chosen distributions, especially if the real distributions are multimodal.

Non-Parametric methods do not rely upon any underlying distributions but instead attempt to estimate the class conditional distributions with an appropriate technique. Non-parametric methods should produce superior classification results when compared to parametric methods, unless the parametric distributions are relevant and are estimated with great accuracy.

Statistical pattern recognition techniques were applied to the data set for the following reasons :

- In order to confirm the non-linear characteristics of the data investigated in section 4.2.3. It was concluded in section 4.2.3 that neural computing techniques would produce superior recognition performance as a result of their ability to model non-linearities more effectively. An inferior performance of

the statistical techniques, compared to neural computing, would strengthen this argument.

- Statistical techniques are well established and consequently results obtained using statistical methodology can be readily compared to previous work.
- Section 4.2.3 concluded that the performance of supervised statistical methods should be investigated following the favourable performance of Linear Discriminant Analysis (LDA).

The statistical techniques were chosen to incrementally support arguments, presented in section 4.2.3, regarding the structure of the data set. Each sub-section will detail the aspect of the data which is under investigation.

5.2.1 Parametric Discriminant Function

This is the only method used in this work which utilised a parametric estimation of the class conditional distributions. The underlying distributions were assumed to be multivariate normal. This technique was specifically chosen to determine whether the data set conformed to a specific statistical distribution, multivariate normality in this case. Following the conclusions made in section 4.2.3 regarding the complex nature of the data it was predicted that this method would produce poor recognition performance which would indicate that the data possessed a complex statistical distribution which required non-parametric estimation.

It can be shown that if a univariate normal distribution is considered then the following discriminant function will minimise the classification error [81]:

$$g_k(\mathbf{x}) = log\, p(\mathbf{x} \mid C_k) + log\, P(C_k) \tag{5.6}$$

Since a total of fifteen features represent each odour sample the discriminant function must be modified in order to apply to a multivariate normal distribution, which is given as follows[81]:

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp[-0.5(x-\mu)^t \sum^{-1}(x-\mu)] \tag{5.7}$$

where

x is the input feature vector.

Σ is the covariance matrix.

μ is the multivariate mean.

Following substitution and simplification the discriminant function for the multivariate case can be expressed in quadratic form as follows [81] :

$$g_k(x) = x' W_k x + w_k' + \omega_{k0} \tag{5.8}$$

where

$$W_k = -\frac{1}{2}\Sigma_k^{-1} \tag{5.9}$$

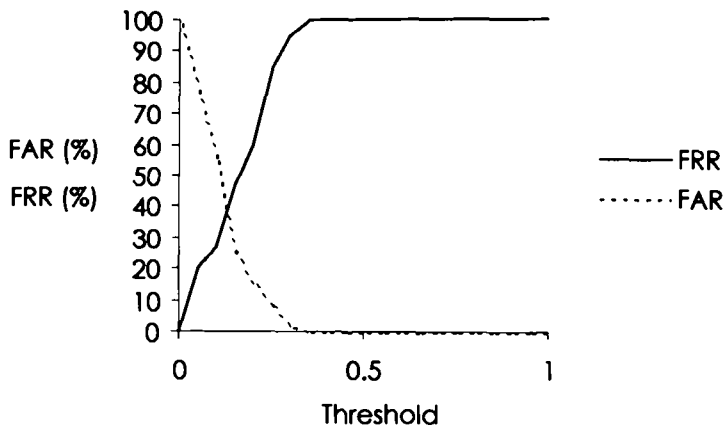
$$w_k = \Sigma_k^{-1}\mu_k \tag{5.10}$$

$$\omega_{k0} = \frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k - \frac{1}{2}\log|\Sigma_k| + \log P(\omega_k) \tag{5.11}$$

The winning class during classification is the class k whose discriminant function, $g_k(\mathbf{x})$, value is the highest.

The resultant FAR/FRR curves, as illustrated in figure 5.3, for this method revealed extremely poor verification performance with a high degree of high probability misclassifications coupled with low probability of correct classifications. Similar results were obtained when biometric recognition was considered with a recognition rate, defined in equation 5.1, of only 34% achievable. Figure 5.20, in section 5.4, shows that this method was actually the poorest classifier for this data set. This poor performance can be attributed to the inaccurate representation of the feature distributions by estimating multivariate normal distributions. Use of non-parametric classifiers should estimate the feature distributions more successfully and hence improve classification performance.

Figure 5.3 FAR/FRR for Parametric Discriminant Function



5.2.2 Non-Parametric Discriminant Function

The discriminant function presented in section 5.2.1 assumed that the odour features were normally distributed. The poor classification proved that this assumption was not applicable for this data set. This section presents a discriminant function which does not rely upon any underlying assumptions about the feature distributions. Although this method is technically non-parametric, it does not actually estimate the class conditional feature distributions but merely attempts to minimise a criterion function. This method was intended as a comparison with the previous method, parametric discriminant function which was presented in section 5.1.2. This method does not make any assumptions regarding the underlying distribution of the data set and was predicted to produce superior recognition performance when compared to the parametric method. However, this method was not expected to produce acceptable recognition performance due to its inherent linear properties and the difficulty in modelling a complex non-linear data set with linear discriminant functions.

If a multi-class problem is considered, such as the eight classes involved in this work, then c linear discriminant functions can be defined as follows [81] :

$$g_k(x) = W_k^T x + w_{k0} \tag{5.12}$$

where

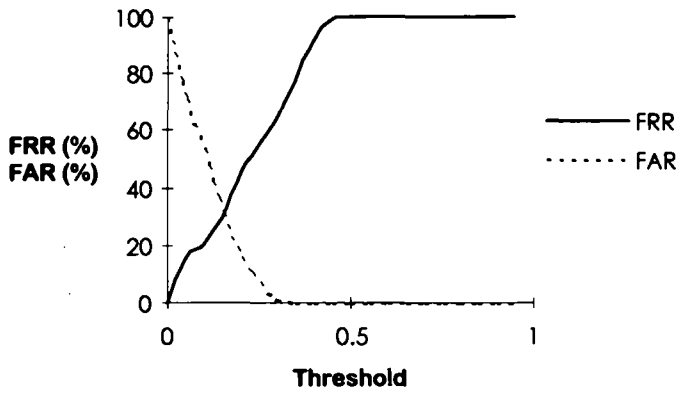
$k = 1,...,c.$
W is the weight vector.
w_{k0} is the threshold weight.

These two weights have the following effects: the normal weight vector determines the orientation of the hyperplane decision boundary, and the location of the surface is determined by the threshold weight w_{k0}. The actual discriminant function $g_k(x)$ is proportional to the signed distance from x to this hyperplane, described by W and w_{k0}. As previously discussed in section 5.2.1, the new pattern x is assigned to the class k which possesses the highest value discriminant function : assign $\mathbf{x}$ to C_k if $g_k(x) > g_j(x)$ for all $j \neq k$. The discriminant function divides the feature space into c regions where $g_k(x)$ will possess the largest discriminant if x is in the region R_k.

When using linear discriminant functions the training data is needed in order to establish the weight vectors which minimise the probability of misclassification. One such method is to minimise the sum of squared errors of the discriminant functions in order to determine the weight vectors. This method uses the error calculated from all of the samples and not only the misclassified samples, which is the case with some error reduction algorithms.

The resultant FAR/FRR curves for this method are shown in figure 5.4, which indicate very poor biometric verification performance. Initially the FAR curve indicates good performance since there are zero false acceptances above a threshold of 0.35. However, this result is ambiguous since this classifier did not actually produce any high confidence matches, illustrated by no correct classifications above a threshold of 0.45. Therefore, this classifier was unable to partition the non-linearities in the feature space resulting in a poor biometric recognition performance of only 54% coupled with very low confidence decisions.

Figure 5.4 FAR/FRR for Non-Parametric Discriminant Function



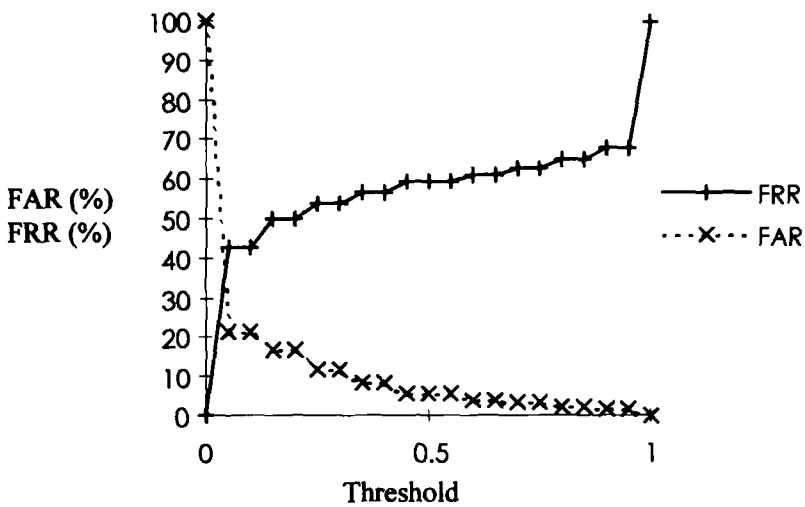
5.2.3 Nearest Neighbour Classification

The nearest neighbour method of classification assigns the new pattern **x** to the class of the training pattern which is closest to the new pattern. This rule can be extended to incorporate k neighbouring patterns. In this case the new pattern **x** is assigned to the class most frequently represented in k nearest samples, contained in the reference data set [79]. This method is a well documented pattern recognition technique[79, 88] and was chosen for use as a standard methodology which the neural computing techniques could be compared against. It was predicted this method would produce superior results to the discriminant function methods, as presented in sections 5.2.2 and 5.2.3 respectively, but would still not produce satisfactory recognition performance as a result of its linear nature.

The classification error can be shown to decrease as the number of nearest neighbours k increases, reaching a theoretically optimal value when k reaches infinity. However, in real problems there is a finite data set so the actual number of neighbours k is limited to ensure that both a reliable estimate is produced and also to ensure that all of the k nearest neighbours are actually very near to the new pattern **x**. Therefore, a compromise is needed which is usually a small proportion of the number of reference samples. This method of classification does not require any induction phase since the training exemplars are used directly to determine the k nearest neighbours.

The distance measure chosen to calculate the distance between the new pattern and the reference patterns is crucial to ensure reliable classification; in this research the Euclidean distance was used. The number of neighbours *k* was varied in order to find the optimal value for this particular data set. The resultant FAR/FRR curves for this method are shown in figure 5.5, which again indicate very poor biometric verification performance.

Figure 5.5 FAR/FRR Curves for k-Nearest Neighbour (k=10)



The FAR curve reveals adequate performance; for example, if a threshold of 0.5 was selected a FAR of 6% was produced. However, the corresponding FRR at this threshold was extremely high at 57%, which indicated that a very high proportion of the new classes were misclassified. This poor performance was confirmed with a low recognition rate of 54%. However, unlike the non-parametric discriminant function this method produced very high probabilities associated with correctly recognised classes. Figure 5.20, at the end of this chapter, reveals that this classifier still produced a 40% recognition rate at a threshold of 0.95 and was the best performing classifier at this threshold.

5.2.4 Non-Parametric Estimation using a Parzen Window

The Parzen Window technique is a non-parametric method which estimates the class conditional distributions by the summation of individual distribution functions placed at training point locations[75]. This statistical method estimates

the underlying data distribution more accurately than the previous statistical techniques and was chosen to confirm the complex structure of the data. Consequently, this method was predicted to out-perform the previous statistical methods since it was able to produce more accurate estimates of class conditional distributions. However, it was predicted that this method would produce inferior results to the neural network techniques due to the assignment of linear decision boundaries.

The estimation of class conditional density function is given by[75] :

$$g(x) = \frac{1}{Q\sigma}\sum_{i=1}^{Q} W\left(\frac{x - x_i}{\sigma}\right) \tag{5.13}$$

where

Q is the sample size.

W is the window function.

σ is the window width.

During Classification the unweighted summation of all kernels belonging to each class is commuted and is divided by the total summation of all kernel functions for all classes[75]. This operation gives the estimated class dependent *a posteori* probability of class membership for each class which can be used directly to determine the winning class.

The choice of window function, W, and corresponding window width, are crucial since they directly effect the summation of the density function, σ, and hence the accuracy of the underlying data distribution estimate[75].

The Parzen Classifier was induced with several different kernels: Gaussian, Hyperspheric and Lorentz. The optimum classification results were heuristically determined for each kernel type by adjusting the window width. The effective selection of the window width was crucial to ensure optimum performance. Figure 5.6 shows the effect the Gaussian window width had on the absolute recognition rate of the classifier, when subjected to the test set. The recognition rate can be seen to sharply decrease either side of the optimum width of 13%. The system

performance was reduced if, for example, a window width of 20% was adopted, resulting in a decrease in performance of approximately 25%.

The Parzen Window width effectively determines the extent of dispersion of each class in feature space. For the case of the Gaussian kernel, the width refers to the spread of the Gaussian distribution. A large window width would overestimate the spread of the class data and increase the probability of error due to significantly overlapping class distributions. Conversely a very small window width would only partially represent the spread of the class data in feature space, again causing a reduction in recognition performance. As previously mentioned, these two extremes can be easily visualised in figure 5.6 below.

Figure 5.6 Effect of Parzen Window Width upon Recognition Rate for a Gaussian Kernel

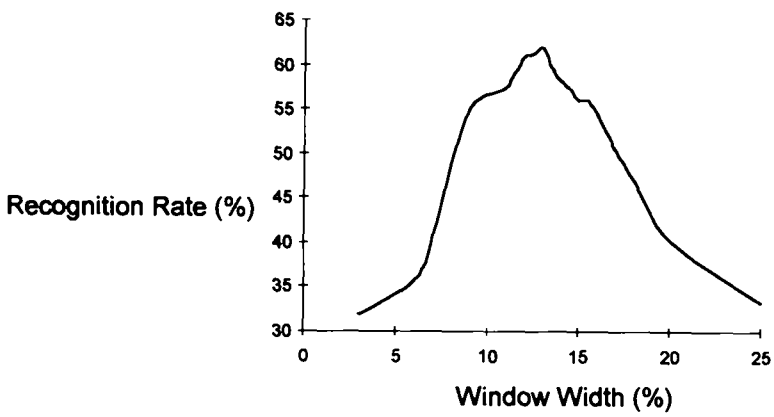
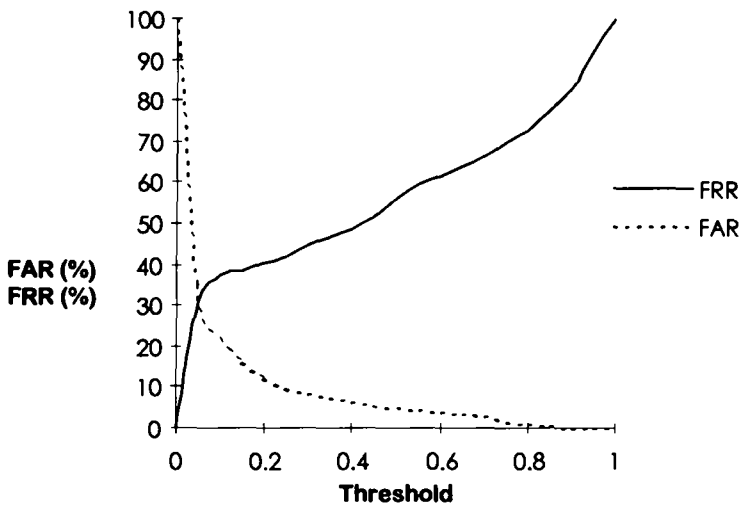


Figure 5.7 shows the resultant FAR and FRR curves, produced using a window width of 13%, for the resulting 'test set' class membership probabilities. The graph illustrates a sharply decreasing FAR curve indicating that the induced classifier produced high posterior probabilities for the winning class for each presented test sample. This attribute is, however, slightly misleading because over 35% of these high output probabilities were, in fact, incorrectly assigned. When averaged over the complete range of output probabilities these incorrectly assigned classes were not adequately represented on the FAR curve due to the high activation levels of the incorrectly assigned classes. These high activation

levels can be seen on figure 5.7 as the FAR curve remains non-zero until a threshold level of 0.9.

The FRR curve, illustrated in figure 5.7, again shows extreme non-linearities caused by poor representation for certain classes. The FRR curve possesses a steep gradient in the threshold region 0 to 0.1 which corresponds to an increase in FRR from 0 to 35%. This increase signifies that 35% of the expected winning classes failed to generate output probabilities greater than 0.1. This observation is in accordance with the FAR comments. The remainder of the FRR curve gradually progresses to the maximum implying that the remaining classes possessed relatively high output levels.

Figure 5.7 FAR/FRR Curves for 13% Parzen Window Width Gaussian Kernel



When the Parzen classifier was considered as a biometric recognition device then an absolute recognition rate, hit rate, of 62% was achieved. Although this rate is relatively poor when compared to other commercially biometric access control systems it is significantly more successful than the previous statistical methods utilised up to this point. The biometric *recognition* performance is compared to other pattern recognition methods in section 5.4, and is illustrated in figure 5.20.

5.3 Neural Computing Techniques

Neural computing is one of the most rapidly expanding areas of current research. Artificial neural networks are simple models of three of the basic properties of biological information processing: parallel operation, simple processing units and synaptic information storage. The basic concept is derived from a model of the biological neuron [89].

Many different algorithms exist, each serving a different purpose depending upon the nature of the problem under consideration. Neural networks can be broadly split into two main groups: supervised and unsupervised networks. Supervised networks involve a training phase which uses class membership information to optimise a network so that subsequent occurrences of similar data sets can be recognised. Unsupervised networks are generally used when the structure of data is analysed since class membership information is not used. One such supervised method is the Kohonen Self Organising Map, as used in section 4.2.3.4 for exploratory data analysis.

Neural Computing methods were chosen for pattern recognition as a result of the conclusions made in section 4.2.3. It was concluded that the 'field trial' data was non-linear in nature which would benefit from the application of neural networks to improve class separability. An unsupervised neural network used, presented in section 4.2.3, produced superior class separation and hence supported this argument. Each sub-section details the reasons behind the choice of the specific neural computing architectures used in this work.

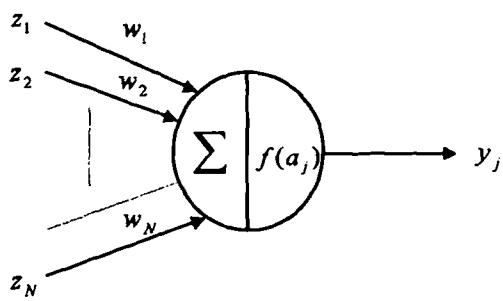
5.3.1 Feed Forward Multi-Layer Perceptron Network

The feed forward Multi-Layer Perceptron (MLP) is the classical neural network architecture and has been subjected to much research and has also been applied to many real world problems. The MLP was chosen for this work since it represents a bench-mark performance from which other neural computing techniques can be compared. The MLP architecture has also proven recognition performance with a

wide variety of different sets. This flexibility is primarily due to the numerous algorithms and control parameters which have been developed to tailor the network to specific data sets and requirements.

A neural network is constructed from many interconnected artificial neurons. The artificial neuron is a simple model of a biological neuron and can be seen in figure 5.8. A single artificial neuron is of little use by itself but when many are interconnected they are able to solve complicated non-linear problems.

Figure 5.8 An Artificial Neuron



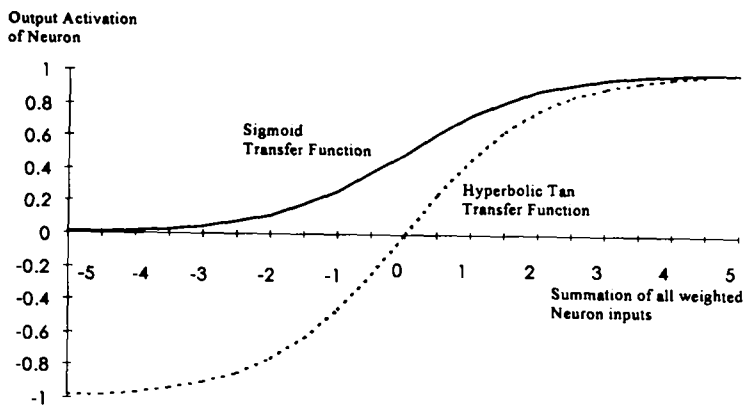
The operation of a single neuron is straightforward and can be summarised into three main steps as follows [76]:

1. Each input to the neuron is multiplied by a weight which can be excitatory or inhibitory. This process is similar to the synaptic strengths of biological neurons which rely upon chemical stimulation to vary an electrical impulse strength into a neuron. The unweighted inputs to a neuron can be represented by the vector $z_0, z_1, \ldots, z_n$ where there are N inputs to the neuron. The weights for the neuron are represented by $w_0, w_1, \ldots w_n$ indicating a separate weight for each input to the neuron. These weights are the memory of the neural network and are distributed across the entire network structure.

2. The individual weighted inputs are summed and are represented by a_j for the jth neuron, which can be compared to the Hillock zone of a biological Neuron.

3. The output activation is then calculated following summation of the weighted inputs. This activation is calculated using a transfer function, $f(a_j)$, which is selected depending upon the network in question. Two common transfer functions are shown in figure 5.9, both provide a graded output which is dependent upon the excitation of the neuron. This graded output is needed so that learning can be introduced into the network.

Figure 5.9 Two Common Transfer Functions : Sigmoid and Hyperbolic Tan



The output from a single neuron can be represented by the following equation [89]:

$$y_j = f\left(w_{ji} z_i\right) \tag{5.14}$$

where

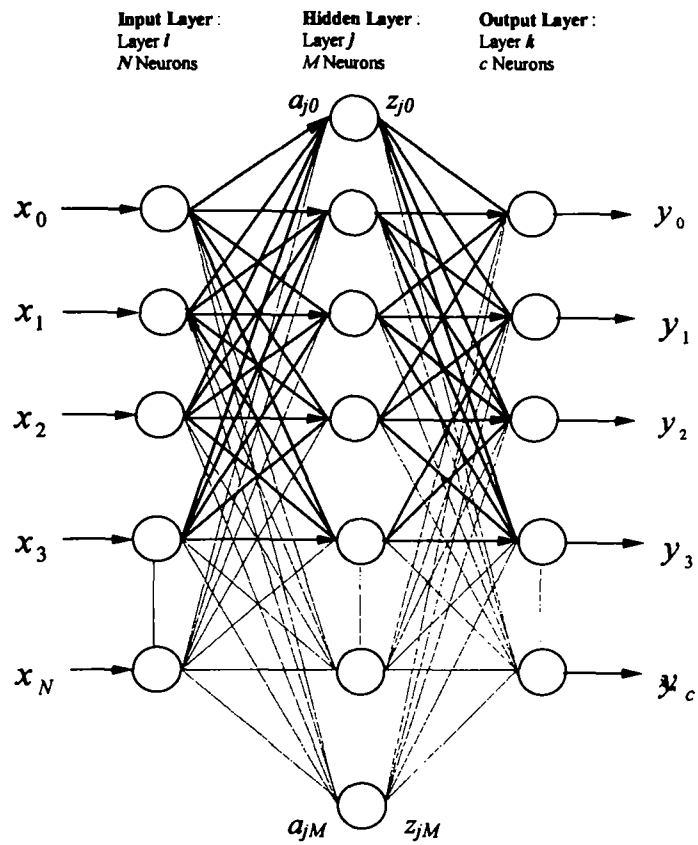
i represents the ith input to the jth neuron.

$w_{ji} z_i$ is the summed input to the jth neuron, represented by a_j.

f is a transfer function applied to the summed input, a_j, to neuron j.

An interconnected multi-layer network is illustrated in figure 5.10.

Figure 5.10 A Multi-layer Feedforward Neural Network



Equation 5.14, for a single neuron, can be expanded to represent the output from an interconnected network of neurons incorporating multiple layer, as illustrated in figure 5.10. The output from the hidden layer of neurons, layer j, can be represented as [87]:

$$z_j = g\left(\sum_{i=0}^{N} a_j\right) \tag{5.15}$$

where

z_j represents the output from neuron j from a total of M, in the hidden layer.

g_j represents the transfer function applied the summed input to neuron j.

a_j is the summed input to the hidden layer neuron j given by :

$$a_j = \sum_{i=1}^{N} w_{ji} x_i \tag{5.16}$$

where

x_i represents the input i to neuron j from a total of N inputs.

w_{ij} represents the weight applied to input i at neuron j.

The outputs from the hidden neurons are then fed forward as the inputs to the next layer of neurons, the output layer in the case of figure 5.10. Networks consisting of two layers of weights are usually sufficient for most classification problems [87] but in some cases a further layer of weights is needed. The neuron activations are fed forward again when a network consisting of three layers of weights is considered. The output activation at neuron k, y_k, for a two layer network can be expressed as [87]:

$$y_k = \tilde{g}\left(\sum_{j=0}^{M} w_{kj}\, g\left(\sum_{i=0}^{N} a_j \right) \right) \tag{5.17}$$

where

$\tilde{g}$ represents the activation function at the output layer.

g represents the activation function at the hidden layer.

w_{kj} represents the weights applied between the hidden layer, j, and the output layer, k.

The network learns the training data patterns by back propagating the error of the network. Training data is the set of known data and associated class information. In the case of the odour sensing array, for example, the training set consisted of pre-processed features for each channel which were used as inputs to the network. The actual user number was also stored so that an assessment could be made of the performance of the network.

Initially the interconnection weights are randomised and the training data is presented to the network. In this work the output of the network was then

compared to the coded user number associated with the data. The user number was encoded using 'one of *n*' coding which means that there were *n* output neurons for *n* classes. The user *n* was represented by output node *n* having a maximum excitation and all other nodes having minimum excitation. Sometimes an extra neuron is added to the output layer, $n + 1$ neurons, which can be used to negatively train the network. Negative training improves novel data rejection, for example an intruder, and involves assigning known null data patterns to the extra *reject* neuron in the output layer. This methodology may increase the intruder rejection performance of an biometric access control device.

The output from the network is then compared to the required output by using an appropriate error function. This error is then back propagated to all previous neurons, or nodes, in the network in order to adjust the weights for optimisation. The contribution to the overall error of a particular neuron is proportional to the weight attached to its output. Hence, the back propagated error to a neuron is scaled by its output weight. The weights are altered according to an error optimisation function such as gradient descent.

The network is trained by applying labelled training patterns to the inputs and then adapting the network accordingly. The network performs this by firstly assessing the output error, E, associated with a particular training example, T, and then by adapting the weights in order to reduce this error. This two staged process is repeated for all of the training exemplars until predefined convergence criterion are fulfilled, for example until the error is reduced to a satisfactory level or a pre-determined number of iterations is reached. This training phase therefore attempts to minimise the error of the network. Convergence of the error function can be extremely difficult due to high dimensioned error mappings which could contain numerous local minima. The choice of error function and error minimisation function are therefore critical to the successful operation of the network.

A commonly used error function is the 'sum of squares' error. The networks weights are therefore adjusted to minimise the total sum of squares error, E, which can be expressed using the Euclidean distance by [87]:

$$E = \frac{1}{2}\sum_{q=1}^{Q}(T^q - Y^q)^2 \tag{5.18}$$

where

q is the current training example from a total of Q exemplars.

T is the target output vector.

Y is the actual output vector produced from training example q.

This error function can be perceived as a function of both the network weights and the input training vectors. Substituting into equations 5.15 to 5.17 this error function can be represented for a single hidden layer network, as shown in figure 5.10, by[87] :

$$E = \frac{1}{2}\sum_{q=1}^{Q}\left(\sum_{k=1}^{c} t_k^q - \tilde{g}(a_k)\right)^2 \tag{5.19}$$

where

q represents the training sample from a total of Q training samples.

a_k is the summed input to the output layer neuron k given by :

$$a_k = \sum_{j=1}^{M} w_{kj} z_j \tag{5.20}$$

where

z_j is the output from the hidden layer neuron j defined in equation 5.15.

This error function can not be solved in closed form but can be approximated iteratively with gradient descent, a learning rule methodology [87]. The local minimum for a non-linear real valued function can be found by setting $dE/dw = 0$ and then solving for w. Since the minimisation is performed iteratively an initial

starting point, w_0, is chosen from which the function is subjected to a step change in the direction of steepest descent. This step change can be expressed for step $r+1$ as :

$$w^{r+1} = w^r - \Delta w^r \tag{5.21}$$

where

Δw^r is the change in network weights as a result of back propagation of error and can be defined as :

$$\Delta w^r = -\eta \nabla E\,|_{w^r} \tag{5.22}$$

where

η is the learning rate, $0 \leq \eta \leq 1$.

$\nabla E|_w^r$ is the gradient of the error in weight space following back propagation of the network error.

The parameter η is the step gain, or learning rate, which amplifies or attenuates the step size [90]. The correct choice of this parameter is crucial to the process since if η is too large then the function would move past the local minimum whilst a low η would require a large number of iterations for convergence or may not reach the local minimum at all. The learning rate is sometimes modified as training is performed in order to encourage convergence. The gradient descent method for error minimisation is extremely inefficient when highly non-linear problems are considered. Many modifications have been derived to increase the performance of this method and are discussed in section 5.3.1.1.

A multi-layer neural network can be trained via gradient descent by back propagating the output error to the previous layers. A neural network contains weights for each layer of the network and these are used to calculate the contribution to the overall error of the network. The change in weights in the second layer of weights can be defined as [87] :

$$\Delta w_{kj} = -\eta.\tilde{g}'(a_k)\frac{\partial E}{\partial y_k}.g(a_j) \tag{5.23}$$

where

a_k is the summation at output neuron k defined in equation 5.20.

$\tilde{g}'$ is derivative of the output layer transfer function.

$\frac{\partial E}{\partial y_k}$ is the derivative of the error function with respect to the output activations. If the sum of squares error function is used, as defined in equation 5.18, this derivative is given by :

$$\frac{\partial E}{\partial y_k} = y_k - t_k \tag{5.24}$$

The change in weights in the first layer in calculated by back propagating the network error and is given by :

$$\Delta w_{ji} = -\eta.x_i.g'(a_j)\sum_{k=0}^{c}\left(w_{kj}.\tilde{g}'(a_k)\frac{\partial E}{\partial y_k}\right)$$

(5.25)

where

a_j is the summation at hidden neuron j defined in equation 5.16.

g' is the derivative of the hidden layer transfer function.

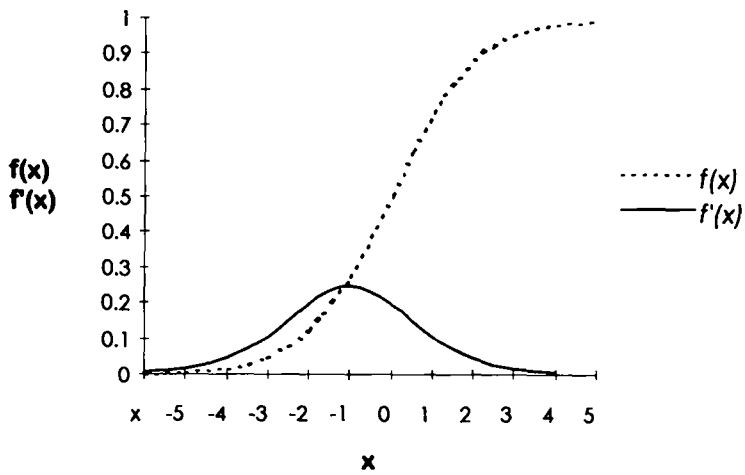
Derivation of the transfer function encourages weights which are dithering and hence attempts to force an output decision. This can be visualised in figure 5.11 which shows the sigmoid transfer function and its associated first derivative. Neuron activations which lie towards the midpoint of the sigmoid curve have consequently high derivatives which produce a high weight modification. Neuron activations which lie towards either extreme of the sigmoid function are not

modified to a great extent due the small derivative value. The Sigmoid transfer function and its derivative can be expressed as :

$$g(x) = 1/1 + e^{-x} \tag{5.30}$$

$$g'(x) = g(x).(1 - g(x)) \tag{5.31}$$

Figure 5.11 The Sigmoid Transfer Function and its Derivative



Too much training can be disadvantageous as the network learns the idiosyncrasies of the data as opposed to the general patterns. The result of this is that the network may recognise all of the learning set with excellent results but when offered with the test set the network will not generalise. A network which is trying to learn minute variations will consequently take a considerable amount of time to learn. The best way to eliminate this problem is to reduce the learning capability of the network by reducing the number of hidden layers or reducing the number of nodes in the hidden layer. The network will then be incapable of learning minute variations and, therefore, will only generalise which is the aim of network for pattern recognition purposes.

An extra term, called momentum, is often added to the gradient descent algorithm in order to avoid oscillations over the error surface which hinders network convergence [87]. Momentum effectively adds inertia to the motion through

weight space and smoothes out oscillations. The weight modification equation defined in 5.22 can be modified to incorporate momentum as follows :

$$\Delta w^r = -\eta \nabla E \mid_{w^r} + \mu \Delta w^{r-1} \tag{5.32}$$

where

μ is the momentum, $0 \leq \mu \leq 1$.

If the error curvature is low then momentum increases effective learning rate and hence speeds up convergence. However, if the error curvature is high then successive momentum terms tend to cancel each other out effectively causing minimal effect to the learning rate. Hence, momentum term can lead to faster convergence without causing divergent oscillations.

5.3.1.1 Network Design and Optimisation

As previously stated a neural network must be correctly designed and optimised in order to extract the true potential of this technology. A poorly designed and implemented neural network may produce extremely inferior results. However, the design and optimisation of a network is time consuming and complex since a great number of parameters and methodologies significantly alter the characteristics of the network. Neural networks must be designed to match the data set in question. The neural networks used in this work were optimised according to the following criteria :

- Network Architecture : number of input nodes, output nodes, hidden layers and hidden layer nodes.
- Activation function
- Error minimisation method
- Training and Evaluation procedures

The input to the neural network was the 3 dimensional optimised feature vector, derived in section 4.2.1, applied to the data of the 5 sensors selected in section 4.2.2. The data set contained data for 8 classes which represented the data supplied by the 8 reliable field trial participants. One input node was assigned to each feature vector which fixed the number of input nodes to fifteen, $N = 15$. A linear transformation was applied to each individual input feature, n, in order to match the features to the activation function utilised, unipolar sigmoid or bipolar hyperbolic Tan. However, the complete range of the activation function was not used, [0,1] or [-1,+1], because infinite weight values would be required to map feature inputs which lie on the extremities of the activation function [90]. The feature vectors were transformed within the range [0.1,0.9] for a sigmoid activation function using the following equation :

$$y = [0.8(x - x_{n(\min)})] / [x_{n(\max)} - x_{n(\min)}] + 0.1 \qquad (5.33)$$

The number of output nodes needed is some function of the number of classes involved in the classification problem, K=8 in this case corresponding to eight people. Since the number of classes was small the 1-of-n output coding was utilised which required J=K output nodes. If the number of classes were to increase then modifications would be necessary in order to avoid training difficulties. For example, if K is large then the network may set all outputs to zero since this represents an easy way to reduce the training error of the system. This problem could be reduced by either using different output encoding or using an appropriate output activation function such as the softmax function, defined in equation 5.34, which produces output probabilities, which would force the network to produce output activations which sum to one. This problem is not applicable to this data set but would be directly applicable for use in a commercial biometric system where the number of users could be considerably greater than eight.

Both sigmoid and hyperbolic Tan were utilised for the hidden layer activation functions. The choice of hidden layer activation function did not affect the

recognition rate of the network but hyperbolic Tan provided both an increased input range and also eliminated the need for a bias in the hidden layer. The Softmax, or normalised exponential, activation function was used for the output nodes so that network outputs were interpretable as posterior probabilities, which was required for evaluation as a biometric device. The softmax function can be defined as [87] :

$$y_m = \frac{\exp(a_m)}{\sum_{k=1}^{c} \exp(a_k)} \tag{5.34}$$

where

y_m is the output from node m.

a_m is the summed input to output node m.

c is the number of classes and hence output nodes, assuming 1 of n coding.

The softmax function has the properties:

$$0 \leq y_m \leq 1 \text{ and } \sum_{j=1}^{c} y_j = 1$$

The outputs sum to one, which are required for probabilities.

The number of hidden layers in the network was initially fixed to one since a single layer network is capable of partitioning any data set provided that there are enough neurons in the hidden layer, assuming that the data set can actually be partitioned[90]. Numerous methods exist in order to determine the optimum number of neurons in the hidden layer [91,87]. The majority of the methods are functions of the number of input nodes, the number of classes involved and the number of training patterns in the data set. However, it is believed [90] that the number of hidden layer neurons is not purely a function of these parameters but is dependent upon the number of unique patterns which exist in the data set. The number of unique patterns is not necessarily equal to the number of classes since

each class may contain several sub-classes. The optimum number of hidden layers is therefore dependent upon the data set in question.

In order to determine the optimum number of hidden layer neurons the following algorithm was applied to each network configuration :

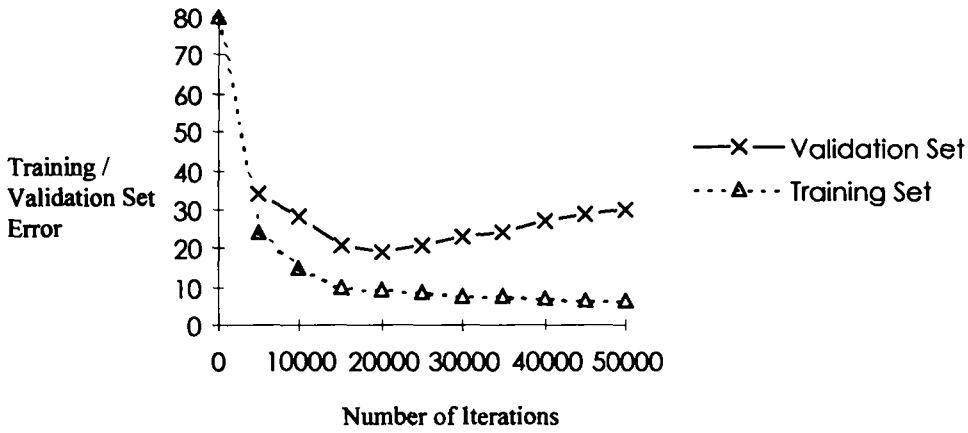
- Start with small number of hidden layer neurons, $k = 2$.
- Train network and assess recognition performance, to both training and validation data, every 5000 iterations in order to determine if overtraining has occurred, see figure 5.12. This process was repeated many times for each architecture in case weight initialisation hindered network convergence.
- Add one neuron to hidden layer and repeat training process until network error ceases to reduce. Following stabilisation of error, increase hidden layer nodes at higher increments, for example following stabilisation at 10 nodes increase nodes to 15, 20, 30, 60 nodes in order to verify node selection.
- Choose minimum number of nodes which ensures minimum network error.

This iterative process resulted in an optimum number of seven hidden-layer neurons. This optimum number applied to almost all of the error minimisation algorithms used with the data set. This optimum number was chosen for several reasons. Firstly, any increase in neurons above seven did not provide an increase in the recognition performance of the network when subjected to the test data. If the number of neurons was increased then the network was easily overtrained which increased the recognition rate on the training data but reduced the recognition rate when the test data was used. However, networks with a greater number of neurons were able to produce similar generalisation capabilities but required greater vigilance when training to avoid overfitting. Networks using only six hidden neurons could also be trained with similar characteristics but training proved far more unreliable; sometimes convergence could not be reached.

In order to prevent over fitting, the validation set was used to assess the performance of the network at pre-defined times, in this work the network was tested after every 5000 iterations. Network training was halted when the validation error began to increase, implying over fitting was occurring, and the

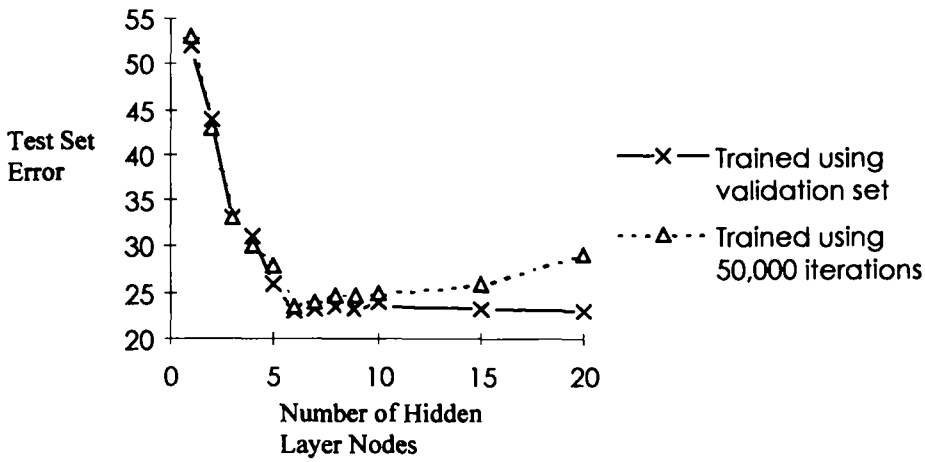
network weights for the lowest validation error were selected. Figure 5.12 illustrates the effect of over training for a network with 20 hidden layer neurons using the 50/10/40 data set. Figure 5.12 shows that even though the training set error reduces consistently as the number of iterations increases, the generalisation ability of the network reduces as can be seen from the increase in error produced for the validation set.

Figure 5.12 Effect of Over Training for 20 Hidden Layer Neurons (50/10/40 Data Set)



The selection of seven hidden layer neurons can be visualised in figure 5.13, which shows the test set error for the 50/10/40 data set, for networks with a hidden layer size ranging from 1 to 20 neurons. The first plot shows the test set error using networks trained by measuring the network generalisation ability by using the validation set, as illustrated in figure 5.12. The test set error can be seen to reduce, as hidden nodes increase form 1 to 7, to approximately 27% (73% absolute recognition rate). The error does not reduce any further as hidden neurons are added hence verifying that 7 neurons contain the required degrees of freedom to represent the data set. The second plot illustrated on figure 5.13 again shows the test set error as a function of hidden layer neurons, but this time the validation set was not used, instead the networks were merely trained for 50,000 iterations. This second plot shows the effect of over fitting since the test set error steadily increases as the number of hidden layer nodes increases past the optimum value of 7.

Figure 5.13 Effect of Hidden layer Neurons upon Training Error



The number of hidden neurons, seven, could imply that one hidden neuron was needed for each unique training pattern [90]. Although eight classes were contained in the data set two classes were responsible for a large degree of the system error, see section 5.3.1.2, which implies there are seven unique patterns. However, although the number of hidden neurons seems to comply with this theory it does not comply with many others which indicates that the hidden layer must be assessed for each data set on an individual basis.

The error minimisation method influenced the convergence of the neural network in the training phase. The standard gradient descent produced satisfactory results by slowly reducing learning rate as convergence approached, termed stochastic approximation. However, initial choice of momentum and learning rate was critical in order to ensure convergence.

Several variations on the standard gradient descent method were utilised which attempt to eliminate the problem of excessive dependence of weight changes upon the magnitude of the gradient. The most successful method was the Quickprop algorithm [92,93] which incorporates numerous heuristics in order to improve gradient descent. These improvements include:

- Modification of derivative calculations to speed up gradient descent.
- A new error function which automatically sets error to zero for small absolute errors which avoids inconsequential weight modification and increases speed.

- Weights are prevented from growing too large by further modification of the gradient calculation.

The Quickprop method provided consistently good results with convergence being reached almost every training attempt, the speed of convergence was also improved. The enhanced performance of the Quickprop has been confirmed by other researchers in the neural network field [93].

The Delta-bar-Delta method also produced satisfactory results and showed higher probabilities of convergence than the standard backpropagation algorithm. The Delta-bar-Delta method monitors sign changes in an exponential averaged gradient and modifies learning rate by the addition of a constant instead of multiplication [92,93].

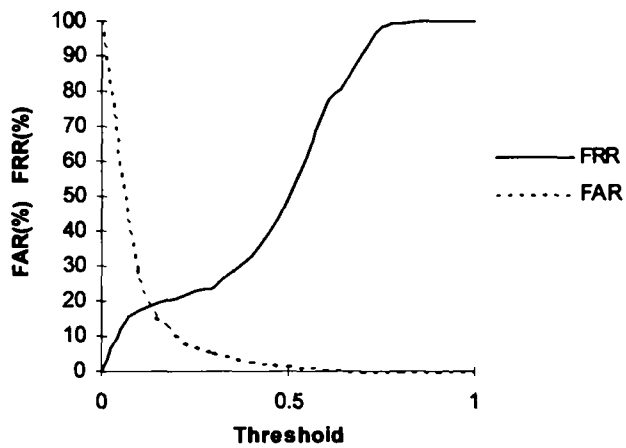
5.3.1.2. MLP Results

The biometric verification results obtained from the optimised network, using 50% of the data for training and 50% for testing, are shown in figure 5.14. Since the Softmax activation function, as defined in equation 5.34, was used at the output layer of the network, then the output activation at each of the eight output nodes corresponded to the probability of membership to the class corresponding to each specific output node. Thresholds were then applied to these class membership probabilities in order to construct the FAR and FRR curves, as described in section 5.1 and as defined in equations 1.11 and 1.12. The FAR curve shows a steep descent indicating that the output activation levels predominantly favoured a single class. Since a high proportion of test samples were recognised correctly then the False Acceptance capability of the remaining class outputs was very low due to the highest probability of the winning class. Figure 5.14 also shows that the FAR curve is almost zero when the output probability threshold reaches 0.5. This indicates that the vast majority of False Acceptances occurred at very low thresholds which shows that when the network was unable to resolve

between several similar classes then low probabilities of class membership were produced.

The FRR curve shown in figure 5.14 still shows non-linearity which is caused by the different recognition rates encountered for each separate class. The FRR curve exhibits a high initial increase in FRR which is mainly due to the samples which were confidently misclassified to the wrong class. This high initial increase is, however, significantly less noticeable than the statistical methods, for example the Parzen window method shown in figure 5.7. The FRR then increases at a reduced rate until higher thresholds of over 0.5 are encountered after which the increase is more rapid. The FRR curve saturates at a threshold of 0.8 which demonstrates that the network did not produce any high confidence matches.

Figure 5.14 FAR/FRR Curves for MLP Network using 50/10/40 Data Set.



When the optimised network architecture was considered as a biometric recognition device, a hit rate of 73% was obtained. Ten networks were trained and tested using the same architecture and the hit rate was obtained by averaging the output from the five best networks[87]. This procedure ensured that poor weight initialisation for a single network did not produce artificially low recognition performance. This network committee procedure also improved network performance[87] by a margin of 0.7% over the best trained single network. This

rate is still unacceptable in terms of a commercial biometric recognition device which exhibit hit rates of over 99%. However, it must be noted that the majority of biometric systems do not actually operate in biometric recognition mode, instead relying upon the relative ease and flexibility of biometric verification. Figure 5.20, at the end of this chapter, compares the hit rates for the various pattern recognition techniques used on the test data. Figure 5.20 shows that this architecture proved to be most effective at discriminating between humans.

Figure 5.15 shows the confusion matrix for the optimised network. The confusion matrix reveals a number of useful properties of the network in question such as individual class hit rates and the allocation of misclassified samples. The diagonal represents the individual class recognition rates, the vertical columns show the percentage of the other classes assigned to the column-heading class and the horizontal rows show the percentage of the row-heading class assigned to the corresponding column-heading class.

Figure 5.15 Confusion Matrix for the Eight People Involved

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	**97.8**	0.0	2.2	0.0	0.0	0.0	0.0	0.0
1	18.4	**0.0**	0.0	7.9	0.0	65.8	7.9	0.0
2	9.3	0.0	**78.7**	0.0	0.0	0.0	7.2	4.8
3	5.5	0.0	1.8	**67.6**	0.0	7.3	1.8	16.1
4	20.0	0.0	0.0	6.0	**58.0**	12.0	2.0	2.0
5	19.3	0.0	2.4	0.0	0.0	**78.3**	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**	0.0
7	5.9	0.0	2.1	3.2	0.0	2.6	0.0	**86.2**

The confusion matrix illustrates that the individual class hit rates varied from either extreme, 0% for class 1 to 100% for class 6, which combined to produce a total hit rate of 73% as previously mentioned. Therefore, the total recognition error of the system can be mainly attributed to class 1 representing 46% of the total system error. If the next worst class is also considered, class 4 with 58% hit rate, then it can shown that these two worst classes were responsible for 73% of the recognition error. This characteristic of a small number of classes being

responsible for the majority of biometric system error has also been observed previously[10].

Further analysis of the confusion matrix shows that the majority, 66%, of the worst class, class 0, was incorrectly allocated to class 5. This was caused by partial overlapping of the two classes in feature space. The neural network training algorithms encouraged a winning class activation when input features were close; in this case the network favoured class 5 but re-training may reverse this allocation due to differing weight initialisation and hence error optimisation characteristics. Analysis of the matrix also shows that class 0 was a popular class for misclassification, a total of 68% of other classes were allocated to class 0. This high rate of misclassifications may have been caused by the large cluster size of class 0 in feature space. The clusters of other classes partially overlapped with the large area of class 0 causing these errors during recognition attempts.

As previously stated this pattern recognition architecture produced the most effective pattern recognition system for this particular data set. This architecture was hence used for more detailed analysis and development in order to more accurately simulate a biometric access control device and also to attempt performance enhancements by use of a time-varying adaptation mechanism, as introduced in section 4.2.3.5.

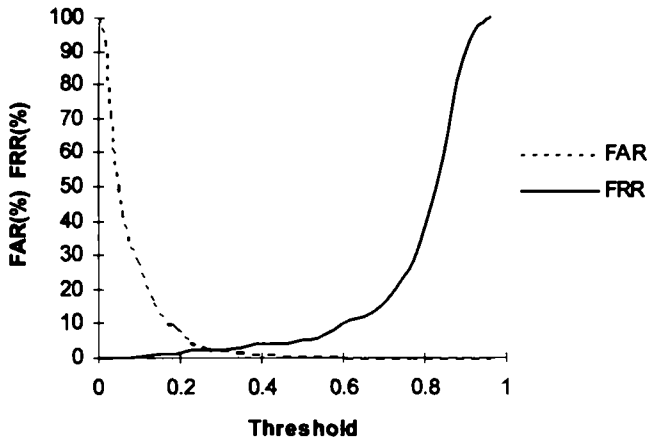
Firstly the data was repartitioned in order to more realistically simulate a biometric access control device. Biometric access control devices usually involve an enrollment session for each system user which gathers a representative number of samples from the user. These samples should ideally represent the unique nature of the individual and also indicate to the pattern recognition system the natural variability of the user's biometric profile. In the case of neural networks this template data can be used directly, following feature extraction, in the training phase. This method of enrollment can be simulated in the data set by using the first two day's data for training and validation, consequently leaving the remaining data for testing. Data for two days was needed to provide an adequate number examples for training; this partitioning corresponded to 10% training data,

5% validation data and 85% test data. This partitioning of the data was a departure from the usual methodology of utilising 60% for training and 40% for test, but was chosen to provide recognition rates similar to those produced by a commercial biometric system. This data set shall be referred to as the 10/5/85 data set.

Figure 5.16 shows the FAR and FFR curves resulting from the 85% test data being applied to optimally configured neural networks; the results were again obtained from a committee of the five best networks. Both curves are similar in appearance to those in figure 5.14, the 50/10/40 data partitioning. However, both curves show more gradual gradients when compared to figure 5.14 indicating that the confidence of the decisions was lower when this reduced training set was considered. The FAR curve illustrates a higher degree of low confidence false acceptances as well as a significant increase in high confidence false acceptances.

The FRR curve also shows a higher rejection rate at lower confidence which again indicates that the network did not predict winning classes with high confidence. This observation is in accordance with those of section 4.2.3.5 where a person's odour profile was shown to vary considerably as time progressed. The hit rate for this architecture was found to be 65%, a reduction of 7.6% from the 50/10/40 data set. This reduction is understandable as the reduced training set was unable to represent the data set in a similar manner due to the time varying aspects of the human odour samples.

Figure 5.17 FAR/FRR Curves for all Eight Classes using Adaptive System over a Two Day Period.



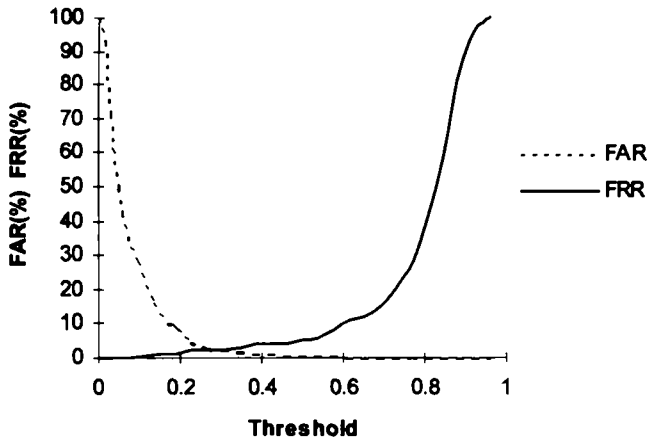
The network was not required to learn such complex decision boundaries since the data from one day only was considered. This enabled the network architecture to be considerably simplified, the optimised network thus requiring only four hidden neurons. The network used ten training samples per user, representing the first day, and then ten test samples per user which represented the following day.

The adaptation algorithm involved re-training the neural network each day to account for the small changes in human odour profiles. The adaptation algorithm effectively tracked the odour profiles of the users as time progressed. The adaptation method operated as follows :

- Train network using ten samples per user for day x
- If a user is correctly recognised on day $x + 1$ then the oldest training example for the recognised user is replaced with the new odour profile template. If a sample is not correctly recognised then the training data set remains unchanged.
- The network is re-trained at the end of the day, ready for the next day.

Therefore, the training data only ever contained ten exemplars for each class, or person. This eliminated various problems associated with increasing the training sample size as time progressed. If the training sample size increased, the complexity of the network had to be modified requiring re-evaluation of the

Figure 5.17 FAR/FRR Curves for all Eight Classes using Adaptive System over a Two Day Period.



The network was not required to learn such complex decision boundaries since the data from one day only was considered. This enabled the network architecture to be considerably simplified, the optimised network thus requiring only four hidden neurons. The network used ten training samples per user, representing the first day, and then ten test samples per user which represented the following day.

The adaptation algorithm involved re-training the neural network each day to account for the small changes in human odour profiles. The adaptation algorithm effectively tracked the odour profiles of the users as time progressed. The adaptation method operated as follows :

- Train network using ten samples per user for day x
- If a user is correctly recognised on day $x + 1$ then the oldest training example for the recognised user is replaced with the new odour profile template. If a sample is not correctly recognised then the training data set remains unchanged.
- The network is re-trained at the end of the day, ready for the next day.

Therefore, the training data only ever contained ten exemplars for each class, or person. This eliminated various problems associated with increasing the training sample size as time progressed. If the training sample size increased, the complexity of the network had to be modified requiring re-evaluation of the

network parameters in order to avoid problems such as undertraining. Discarding old data also improved classification performance by reducing the overlap of classes in feature space. This overlap was reduced because a high proportion of the overlap was not caused by short term, daily, sample deviations but by the similarity between classes measured several days apart. For example, two users may have been separated between days one and two but if the odour sample of user one on day ten was compared to user two on day one then confusion arose.

This adaptation algorithm increased the biometric recognition hit rate from 66%, for a 'real' system trained on the first day's template data only, to 89%. This increase was impressive and progressed towards the level of acceptability for a biometric access control device. The FAR/FRR curve is similar to the two day case shown in figure 5.17 but with reductions in gradient which represented decreases in performance for certain days in the sample period. The average FAR/FFR curve therefore illustrates that the device could operate within acceptable limits if biometric verification was required. A FAR of 0.8% could be obtained by selecting a threshold of 0.5. This FAR is acceptable but it results in a high FRR of 9%, which could increase frustration levels.

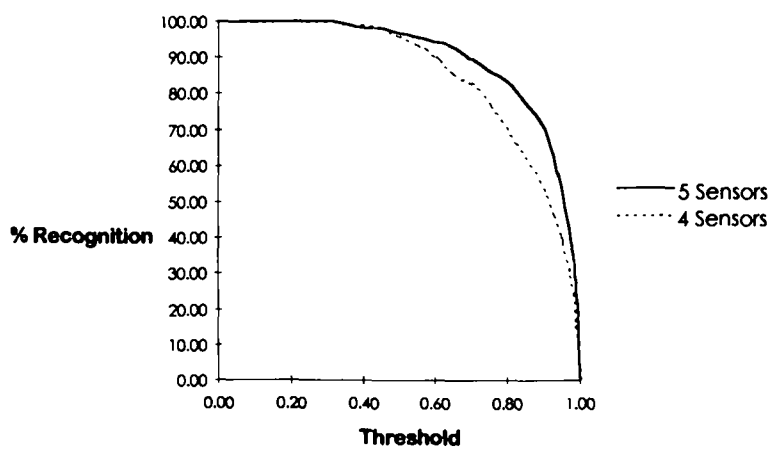
This method of adaptation, however, still suffered from certain drawbacks. The system was still partially reliant upon obtaining representative samples on the enrollment period. If these initial template samples were unreliable then system performance could be significantly reduced. The system performance could not be accurately predicted from day to day because analysis of recognition rates throughout the sampling period indicated that rates varied considerably from as little as 80% to as high as 93%. The lower limits of recognition can be partially attributed to Mondays which suffer from two days delay in adaptation due to the lack of odour samples over the weekend. However, these drawbacks are not specific to this method of adaptation but are applicable to all of the pattern classification methods considered in this section.

5.3.1.4 Effect of Sensor Failure

Since the biometric system was constructed using an array of sensors it is crucial to consider the effect of sensor failure upon the classification performance of the pattern recognition module. Following feature selection in section 4.2.1 the resulting sensor array, used for pattern recognition, consisted of only five sensors, reduced from sixteen. Analysis of previous conducting polymer sensor failures indicated that failure almost exclusively resulted in a reduction in electrical resistance. This decrease in resistance moved the failed sensor out of the dynamic range of the data acquisition system. The features measured for a failed sensor therefore resulted in null readings as there were no measurable deviations from the baseline, due to the resistance shifting out of the range of the data acquisition system.

In order to simulate sensor failure the three input features, corresponding to the failed sensor, were set to zero for all samples in the test data. The test data was then applied to the network. Figure 5.18 shows the decrease in performance when a sensor failure occurred. The recognition rate was significantly lower at certain thresholds than the fully operational sensor array. The worst degradation was a reduction of 10%, observed when the threshold was 0.8.

Figure 5.18 Effect of Sensor Failure upon Recognition Rate using BP Network



This reduction in performance still, however, surpassed the best statistical classifier operating at full sensor capacity. This ability is attributed to the generalisation capabilities of a successfully trained neural network. If the network was overtrained the degradation in performance due to sensor failure would be far greater than observed in this case. It must therefore be re-emphasised that the neural network must be trained effectively in order to ensure the robustness for which neural networks are so renowned.

5.3.2 Radial Basis Function Network

The Radial Basis Function (RBF) network is a relatively new type of feed forward network which differs greatly to an MLP in both its choice of activation function and how it is used. This neural network architecture was chosen for application in this work due to its improved novel data rejection capabilities. Novel data can be defined as data which is not represented by the training data set distributions. When considering an access control system, this novel data represents intruders so an architecture which rejects a higher proportion of intruders is essential. The RBF network also provides greatly decreased training times when compared to the MLP. Quick training is also essential when considering an access control device which must consistently adapt as new users are added.

The RBF network consists of the following three distinct parts [90]:

- An input layer consisting of a node for each feature vector, each of which branches off and interconnects to the following layer, as with the MLP.
- A hidden layer of neurons where each neuron contains a radial symmetric activation function located on the centre of a cluster or sub-cluster in the feature space.
- An output layer of neurons which sum the outputs from the hidden neurons.

The hidden neurons operate in a similar manner to the MLP in that a summation function is performed on the inputs from the feature neurons and in that a transfer function is used to provide an output from the hidden neuron. The difference

between the RBF and the MLP arises in the method of summation and the nature of the transfer function.

The summation function operates in the following manner; each input to a specific hidden neuron has an associated weight, V_{im}, where i is the input neuron number and m is the hidden neuron number. This weight is not multiplying, as is the case with MLPs, but represents a cluster, or sub-cluster, centre[94]. The distance from the current input feature to the stored cluster centre for the input is then computed; this distance is calculated for each individual input to the hidden neuron. A single measure of closeness between the cluster centre and the N dimensional input vector is produced by use of suitable measure, usually the Euclidean distance which is given by :

$$I_m = \sqrt{\sum_{i=1}^{N} (X_i - c_{mi})^2} \tag{5.35}$$

where

I is the Euclidean distance for a specific hidden neuron.

X is the input vector.

c is the cluster centre vector.

This distance measure I is then mapped using a Radial Basis transfer function in order to provide an assessment of membership to the cluster of centre c. This transfer function is often a Gaussian function which outputs stronger values when the distance is small.

The Gaussian function is given by

$$v_m = \exp\left(\frac{I_i - R^2}{\sigma_i^2}\right) \tag{5.36}$$

where

V_m is the output from hidden neuron m.

σ_m is the parameter used to control the spread of the function so that its values decrease more slowly or more rapidly as X moves away from the cluster centre c.

The output neurons from the hidden neurons are finally weighted and summed at the output layer, hence the output layer neurons use a linear activation function. A bias b_j can be applied at each output node to ensure non-zero mean values of the summations. The outputs from the network can be represented as :

$$y_m = w_{1m}v_1 + \cdots + w_{lm}v_l + b_m \tag{5.37}$$

The Training Algorithm of the RBF network is needed to provide the following:

- determination of the cluster centre vectors C_n which is performed by a suitable clustering algorithm, adaptive k-means in this case[94].
- determination of the parameter σ_k which is responsible for the spread of the Gaussian Transfer function for hidden node m.
- construction of an initial weight set for the output layer of neurons.
- perform supervised training of the weights W_{im} in the output layer in order to minimise the total error of the network. The error of the network can be represented by:

$$E = \sum_{q=1}^{Q}\sum_{m=1}^{M}\left(y_m^{(q)} - t_m^{(q)}\right)^2 \tag{5.38}$$

where

q is a training example from a total Q examples.

m is an output node from a total of M.

y is the output from a node m.

t is the ideal target at node m.

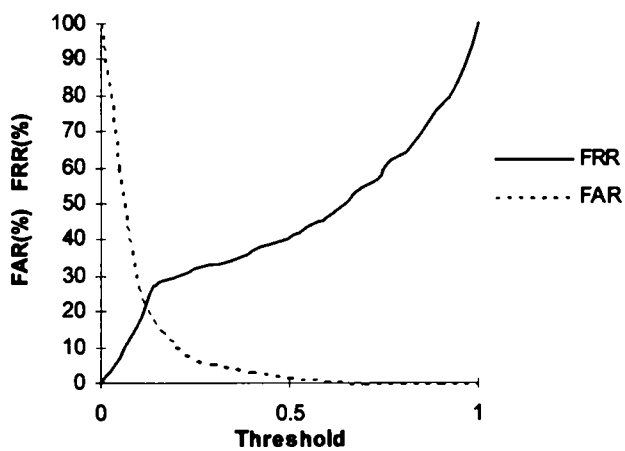
The 50/10/40 data set was applied to the RBF network and the optimum number of hidden nodes was determined heuristically, in a similar manner to the algorithm

presented in section 5.3.1.1. The optimum architecture was found to be 15 input nodes, 130 hidden nodes and 8 output nodes utilising 1 of n encoding.

The FAR and FRR curves were constructed in an identical manner to the MLP results, as presented in section 5.3.1.2. This was possible due to the use again of the Softmax activation function, as defined in equation 5.34, at the output layer of the network. The output activation at each of the eight output nodes then corresponded to the probability of membership to the class corresponding to each specific output node. Thresholds were then applied to these class membership probabilities in order to construct the FAR and FRR curves, as described in section 5.1 and as defined in equations 1.11 and 1.12.

The results using the RBF network were disappointing as figure 5.19 reveals. However, the FRR curve reveals a high proportion of high confidence matches since the FRR curve does not reach 100% until a threshold of almost one. This high confidence for matches was far superior to the back propagation method, as discussed in section 5.3.1, but had the consequence of a lower recognition rate of 65%. This method is compared to the other pattern classifiers in figure 5.20 at the end of this chapter.

Figure 5.19 FAR/FRR Curves for RBF Network using 50/10/40 Data Set.



This neural network architecture may have produced inferior results due to both the large number of input features and the large number of training exemplars,

since this method is more suited to small data sets [94]. RBF networks are suited to more compact applications. Another possible reason for poor performance is the inability of the primary cluster phase to partition the data effectively. This ineffectiveness could have also overestimated the spread of the Gaussian functions in the hidden layer.

The RBF network also produced inferior intruder rejection results when compared to the MLP. It was hoped that the RBF network would have produced better results since the Gaussian functions produce a low output if the input signal is out of range, as opposed to the MLP which has infinite hyperplanes. The overestimation of the Gaussian spread would again explain the poor intruder rejection characteristics of this network.

5.3.3 Learning Vector Quantisation Network

The Learning Vector Quantisation (LVQ) network is an adaptation of the Kohonen Self Organising Map (SOM) network, as described in section 4.2.2.4. This neural network architecture was chosen since the SOM produced good separability between classes during exploratory data analysis, as described in section 4.2.2.4. The SOM was, therefore, predicted to produce similar class separability when used within a supervised pattern recognition architecture. This architecture also differs significantly from the two previous architectures, MLP and RBF, and hence facilitates a comparison of differing neural network architectures. Previously the Kohonen network was used for data exploration as no class information was supplied to the network, the network then determined any relationships with the data supplied. The LVQ adapts the Kohonen network for pattern recognition by adjusting the Kohonen layer depending upon the classification performance of the network.

The LVQ network operates in the following manner. The M neurons of the Kohonen layer are initialised by assigning the first M input feature vectors of the training set to each of the neurons successively. The number of neurons, M, is a multiple of the number of classes involved in the classification problem.

The winning Kohonen neuron is determined by calculating the Euclidean distance, d_i, between the current training vector, x, and the neuron's weight vector, wi. This distance is given by :

$$d_i = \sqrt{\sum_{j=1}^{N}\left(w_{ij} - x_j\right)^2} \tag{5.39}$$

where

N is the dimension of the input feature vector.

The winning distance is the neuron with the smallest distance, d_i, from the training vector, x.

The winning neuron's weight vector is adjusted depending upon the success of the winning neuron as follows [89]:

- if the winning neuron is in the correct class then the neuron weight vector is adjusted to reduce the distance between the input vector, x, and the weight vector, w_i.

  $$w' = w + \alpha(x - w) \tag{5.40}$$

  This process is called reinforcing and the degree to which the neuron weight vector is reinforced is governed by α, where $0 \leq \alpha \leq 1$.

- if the winning neuron is *not* in the correct class then the neuron weight vector is adjusted to increase the distance between the input vector, x, and the weight vector, w_i.

  $$w' = w + \gamma(x - w) \tag{5.41}$$

  This process is called extinguishing and the degree to which the neuron weight vector is extinguished is governed by γ, where $0 \leq \gamma \leq 1$.

This process is repeated for all of the training set exemplars. Following a complete cycle of the training set exemplars the reinforcing/extinguishing factors, α and γ, are gradually reduced.

In this research, once the LVQ network was trained, new input feature vectors were introduced to the network and the winning Kohonen neuron calculated as described above. The disadvantage with this method of class assignment, based upon a winning neuron, was that probability of class membership, or posterior probability, was not provided. The output activation of the winning class was 1 whilst the output of all other classes was 0. Therefore, no information could be determined regarding the confidence of the winning decision or the degree of misclassification[90].

The 50/10/40 data set was applied to the LVQ network and the optimum number of hidden Kohonen nodes was determined heuristically, in a similar manner to the algorithm presented in section 5.3.1.1. The optimum architecture was found to be 15 input nodes, a 10 x 10 Kohonen hidden layer providing 100 Kohonen nodes and 8 output nodes utilising 1 of n encoding.

As previously mentioned the nature of the LVQ output activation was to allocate one winner with a probability of 1 whilst all others had 0 probability. Therefore, the LVQ network was unsuitable for constructing FAR and FRR curves. The LVQ network could be utilised in a biometric verification system but would only perform to the same level as a biometric recognition device. The absolute recognition rate for the 50/10/40 data set, obtained by a committee of the five best LVQ networks, was 72%. This rate was comparable with the other pattern recognition systems as illustrated in figure 5.20. This rate was only slightly lower than the back propagation method but had the associated benefits of this architecture.

5.4 Comparison of Pattern Classifiers

This section compares the various statistical and neural computing techniques used in this work. The methods are compared using the biometric *recognition* performance measures. These measures comprise the absolute recognition rate (ARR) defined in equation 5.1, also known as the hit rate, and the recognition rate (RR) curve as a function of acceptance threshold.

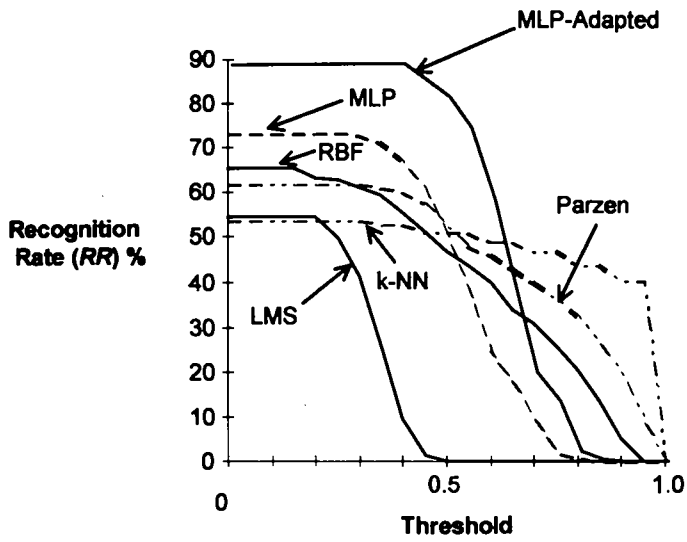
The comparison table as shown in figure 5.20 clearly shows that all of the neural computing techniques outperformed the statistical techniques, when considering the ARR. This superior performance can be attributed to the ability of the neural techniques to model the non-linearities of the data set which were investigated in section 4.2.

Figure 5.20 Absolute Recognition Rate (*ARR*) Comparison for Various Classification Techniques

Classifier	nature	Data Set	
		8 Class : 50% Training	Time Adaptation
LMS	Non Para. : deterministic	55%	
KNN	Non Parametric	54%	
Parzen	Non Parametric	62%	
Quadratic	Parametric : deterministic	34%	
BP	Non linear : neural net	73%	89%
RBF	Non linear : neural net	66%	
LVQ	Non linear : neural net	72%	

The best performing statistical classifier was the Parzen window method which provided a recognition rate just below the RBF neural network. The similarity of these two techniques can be observed in figure 5.21 which shows the recognition rate as a function of threshold. This similarity was most probably due to their common use of the Gaussian function. The increase in performance of the RBF network can be attributed to the non-linearity of the network interconnections.

Figure 5.21 Comparison of Recognition Rate (*RR*) for Various Classification Techniques



The worst classifier was the parametric discriminant function which achieved a recognition rate of 34%. This poor performance can be attributed to both a lack of ability to cope with non-linearities in the data and also the poor parametric approximation of multi-dimension normality.

The most successful methods were the MLP and the LVQ neural networks, both providing a recognition rate of just over 72%. However, in order to produce this high recognition rate the MLP produced lower confidence decisions, as demonstrated in figure 5.21. This low confidence was caused by the inter-class interference caused by the highly non-linear nature of the data and also the time variance of the class data. The feature extraction algorithms, presented in section 3.2, illustrated both the high levels of inter-class class overlap and also the degradation of class separability caused by time variance of the odour responses. The LVQ network uses a Kohonen primary clustering phase which is similar to the SOM used in section 4.2. Therefore, the success of the LVQ consolidates the high discriminatory ability of non-linear linear neural networks, as initially discussed in section 4.2.

The RBF and Parzen window methods provided higher confidence decisions for their correct classifications but suffered from lower overall recognition rates. It must be noted that the LVQ network could not be subjected to a threshold since the output was a binary decision, either 1 or 0, and was hence not represented in figure 5.21.

Figure 5.20 and 5.21 illustrate the performance of the time adapted MLP which produced an average recognition rate of 89% which exceeded all other methods by a significant margin. However, there was still a lack of the high confidence of decisions which was also observed for standard MLP network.

6. Discussion

The problem of biometric identification of humans using human odour has been investigated for application in an access control system. The biometric system was used in the authentication process which transformed a user into the relevant access control system subject. A prototype olfactory device was constructed which sampled the user's odour and then provided the probability that the user was an enrolled subject of the access control system; a threshold was then applied to this probability. This work was reliant upon several new and relatively unexplored areas of research, notably the effectiveness of conducting polymer gas sensors, the complexity of multiple odour diffusion and interactions, and the incomplete work identifying genetically unique human odour descriptors.

Analysis of current access control technologies revealed that the majority of systems used some form of tag issued to users. Many biometric systems, using various different biometric identifiers, had been developed, or were in the development phase. However, although many field trials had been conducted, biometric technology had not become widely used in access control markets. This slow acceptance of biometric technology was attributed to factors such as poor reliability, intrusive user interfaces, lack of a standard biometric feature and public disdain for biometric technology. The use of biometric technology was, however, increasing and it was predicted that it would eventually see widespread use. It was predicted that the use of human odour could provide the ideal biometric system since a user's unique odour could be detected without the knowledge of the user.

References were scarce in the areas of human odour analysis and sensing. Searches for research regarding human odour based biometric access control systems proved unsuccessful and following discussions with biometric experts, it was believed that this work was one of the first of this kind. This lack of research may have been caused by the unique nature of the work and the reluctance of institutions to publish work as a result of commercial constraints. Although biometric electronic noses for access control were scarce, some work had been undertaken in medicine to identify human

odours responsible for various diseases such as pneumonia, infected wounds and lung infections.

Analysis revealed that odour diffusion varied temporally and spatially from an odour source. A human body was regarded as a replenished odour source, in which emitted odour molecules were replaced as they diffused into the atmosphere. However, variations in odour concentration may have occurred due to biological factors. Odour concentration was shown to reduce quickly as distance from the source increased which confirmed odour diffusion models investigated in section 3.2. Close contact, less than 1cm, with the sensors was necessary in order to provide sufficient concentration to prompt a discernible sensor response. The electrical gain applied to the sensor signals was increased in an effort to increase the detection distance but high noise levels restricted gains to levels below 3. It was predicted that odour concentration ratios, even between similar molecular weight compounds, would vary considerably as the distance increased from the source, indicating that an odour 'smells' differently depending upon the distance of the source from the sensors. Sampling as close to the odour source as possible increased concentration and also provided a more representative odour composition since the odour compounds with differing molecular weights had not diffused to different levels.

The chosen sampling system involved static sampling of the hand where the skin of the hand was placed at a constant close distance to the sensors to maximise concentration and also to ensure distance from the source did not fluctuate between samples which would cause error. A dynamic sampling system proved unsuccessful because the source odour was too diluted within the flow chamber to detect a sensor response. The sampling system was ergonomically pleasing and proved to be easy to use which decreased frustration levels and increased the user's perception of the biometric device

The majority of published work regarding electronic nose pattern recognition utilised only maximum divergence responses from the sensors and consequently discarded other sample characteristics of high information content. In this study many other features were extracted from the sensor signals so that the discriminatory contribution

of each could be assessed. Adsorption and desorption characteristics caused by odour-sensor interaction were predicted to be particularly useful for discrimination. A set of 16 different features was presented and was calculated for every sensor and every sample. These features effectively reduced the sample data size from 370 data points for each of the sixteen sensors to 16 one dimensional feature attributes for each of the 16 sensors. Feature analysis revealed that the three features of normalised maximum divergence, normalised adsorption gradient and normalised desorption gradient provided information with the highest degree of differing discrimination characteristics. Other features were discarded due to lack of discrimination and very high correlations with the three chosen features, which indicated similar discriminating characteristics and data redundancy. The addition of any of the remaining 13 features would have merely duplicated discriminatory information or provided redundant information which could have reduced the performance of subsequent pattern recognition techniques. Normalised features across the entire array provided superior discriminatory capabilities compared to features calculated from individually standardised sensor data sets. The standardising methods emphasized the contribution of poorly performing sensors, and consequently de-emphasized sensors showing good discrimination. Normalising reduced the concentration effects of a sample but retained relative sensor response levels.

This optimised feature set, 3 feature attributes per sensor per sample, was then used to assess the individual discriminatory contribution of each sensor in the 16 dimensional sensor array. Analysis showed that two sensors suffered complete failure and hence did not contribute any discriminatory information. A further two sensors responded well to human odour compounds but did not contribute any information which would have aided in the discrimination between the people involved in the field trial; this was due to similar responses obtained for all people involved in the field trial. This common response could have been caused by the sensor responding to common human odour compounds instead of unique compounds. A total of twelve sensors were selected which contained discriminatory information. This sensor subset was then reduced further to a subset of five sensors to eliminate redundancy, since seven of the sensors contained similar discriminatory information. Each odour sample was then represented as an array of fifteen one dimensional features, three features

calculated for each of the five selected sensors. This array of fifteen features was then used for all subsequent feature extraction techniques. This optimised feature/sensor subset was also used for all of the pattern recognition techniques applied in chapter 5.

Section 4.2 investigated several feature extraction algorithms in an attempt to determine any characteristics within the data. The fifteen dimensional optimised feature/sensor set was used as the input to all of the algorithms. The data was restricted to two classes, people in this work, to enable easy interpretation in two dimensional plot of the extracted features. The data for the entire field trial period, six weeks, was initially used in the analysis.

Four algorithms were chosen, each intended to illuminate separate characteristics of the data. Principal Components Analysis (PCA) produced poor results which revealed a large degree of overlapping between the two classes. It was deduced that the data contained non-linear relationships which PCA was unable to separate because of its purely linear behaviour. The technique of Sammon mapping (SAM) was then used, which began with a PCA phase but then progressed non-linearly to maximise the class separation. The SAM produced superior results when compared with PCA, which again implied that the data set contained non-linearities. However, despite the SAM's non-linear behaviour a large degree of overlap between the two classes still existed. The Kohonen Self Organising Map (SOM) produced far superior separability when compared to the techniques of PCA and SAM. The SOM was a non-linear technique but differed greatly to the SAM since it possessed a Neural Network architecture. The two classes were clearly visible with low overlap between the classes, however, a non-linear boundary still existed between the two classes. It was deduced from these results that a non-linear pattern recognition method would provide optimum results. The SOM results also suggested that neural computing methods could enhance pattern recognition performance.

The technique of Linear Discriminant Analysis (LDA) was also applied to the optimised data set. This technique differed from PCA, SAM and SOM because it actually used a priori class information to maximally separate the data, whereas the previous three techniques merely searched for patterns within the data without the use of a priori information. Despite the purely linear nature of this technique it produced

two easily definable clusters for the two classes with only minimal overlap between the two classes. Although the success of this technique was attributed to its supervised nature it was deduced that linear pattern recognition techniques could have also produce similar results since they are also supervised in nature.

Further analysis of the SOM data revealed that the two classes were more easily separated on shorter timescales. For example, the two classes, people in this work, were easily separated when a two day period was considered but classes became increasingly overlapped as time increased. This time variance observation was confirmed for all people in the field trial. When all eight people were considered on a single day only then 80% of the classes were separated but separability reduced when data from a longer timescale was considered.

The optimised feature/sensor data was used as the input data for all of the pattern recognition algorithms used in this work. When the biometric odour system was used as a biometric recognition device, using the 50/10/40 data set, the most successful methods were the MLP and the LVQ neural networks, both providing a recognition rate of just over 72%. The LVQ network used a Kohonen primary clustering phase which was similar to the SOM used in section 4.2. Therefore, the success of the LVQ consolidated the high discriminatory ability of non-linear linear neural networks, as initially discussed in section 4.2. Neural computing techniques outperformed the statistical techniques. This superior performance was attributed to the ability of the neural techniques to model the non-linearities of the data set and the ability to model the data distribution more effectively.

The MLP network used a 10-5-85 architecture representing fifteen input neurons, eight hidden neurons and eight output neurons. One input neuron was used for each of the 15 inputs of the optimised feature set. The output layer used 1 of *n* coding hence provision for one output neuron per class was made. For the case of a commercial system, an extra neuron could have been utilised in the output layer so that negative training could be performed which could improve novel data rejection, in this case decrease the FAR of the system. The number of hidden layer neurons was determined using a network growth algorithm which added neurons to the network until no further performance increase was noted. A network pruning algorithm could also have been

used to determine the optimum number of hidden layer neurons and would have produced a similar network architecture. The optimum number of hidden layer neurons was intended to provide sufficient power to discriminate between the eight people in the field trial but also not over-providing power to the network which would have easily caused overtraining.

The error function used for the MLP network was the commonly used sum of squares function but other error functions which were more suited to 1 of *n* coding, for example the cross entropy error function, could have provided faster training and higher levels of network convergence, but were out of the scope of this work. The most successful learning algorithms were Quickprop and Delta-bar-Delta. However, when compared to the standard gradient descent algorithm they merely provided faster and more reliable network convergence; the absolute best network performance was not specific to a particular learning algorithm. All of the learning algorithms applied in this work were derivations upon the basic gradient descent algorithm but other non-gradient descent learning algorithms, for example quasi-Newton and conjugate gradients, may have provided quicker convergence and enhanced decision boundaries. The application of other learning algorithms was out of the scope of this project. The softmax output activation function was applied to the output layer of the network which enabled the outputs to be interpreted as probabilities. This characteristic allowed the comparison of network performance to other networks and also the statistical methods. However, the softmax function would not have been applicable if the number of output neurons was increased significantly since the learning phase could have artificially reduced error by setting all outputs to low probability values.

The best performing statistical classifier was the Parzen window method which provided a recognition rate just below the RBF neural network. This similarity was most probably due to their common use of multiple Gaussian functions to model the underlying distribution of the data. The increase in performance in the RBF network was attributed to the non-linearity of the network interconnections in the second layer of the RBF. The worst classifier was the parametric discriminant function which achieved a recognition rate of only 34%. This poor performance was attributed to both a lack of ability to cope with non-linearities in the data and also the poor parametric

approximation of multi-dimension normality which was used to generate the discriminant functions.

The MLP produced the highest recognition rate but in order to produce this rate, the MLP produced lower confidence decisions. This low confidence was caused by the inter-class interference due to the highly non-linear nature of the data and also the time variance of the class data. The feature extraction algorithms, presented in section 3.2, illustrated both the high levels of inter-class overlap and also the degradation of class separability caused by time variance of the odour responses. The RBF and Parzen window methods provided higher confidence decisions for their correct classifications but suffered from lower overall recognition rates. It must be noted that the LVQ network could not be subjected to a threshold since the output was a binary decision, either 1 or 0, and hence FAR, FRR and RR curves could not be constructed.

The MLP system was used for further analysis since it produced superior results with the 50/10/40 data set and also due to its high degree of flexibility. The recognition rate reduced significantly to 63% when only 10% of the data set, the 10/5/85 data set, was used for training. This configuration was a closer representation of an access control system, where a small initial enrollment session would provide all of the class templates. This recognition rate was too low for a commercial system since rates greater than 99% would be expected. However, a standard method of calculating FAR and FRR had not been established, so comparing different biometric methods, and even different products using the same biometric identifier, was unreliable. For example, manufacturers may not subject their own biometric system with the same data as a competitor. Ideally, manufacturers using a specific biometric identifier should publish their FAR and FRR based upon a standard data set, hence enabling a direct comparison between devices. The establishment of general biometric standards which would enable direct comparison between different biometric identifiers is more complicated. However, standards will eventually be implemented and could involve synthesis of standard data sets which exhibit similar statistical variations, for example equal variance between finger print data and odour data.

The characteristics of the time variance observation were used to develop a method which exploited the low within-class variance over short periods of several days and

involved adapting the odour template for an individual following each successful recognition attempt. This adaptation method increased the overall recognition rate to 89% which began to approach acceptable levels for a biometric device but still required significant performance advancements. Analysis of the confusion matrix between the various classes involved revealed that 73% of the misclassifications were produced by 20% of the people, a result observed in other biometric disciplines. Consequently, this 20% of the people were subjected to the majority of the frustration caused by the system.

The time adapted MLP produced an average recognition rate of 89% which exceeded all other methods by a significant margin. However, the MLP still exhibited a lack of high confidence decisions which was also observed for standard MLP network. Although the effect of time variance could be reduced with the advent of more effective sensors, it is probable that a certain degree of time variance will always exist as a result of shifting human odour patterns caused by biological factors and interferring odours such as petrol. Hence, the time adaptive algorithm could provide beneficial effects regardless of future advancements in odour sensing technology.

The use of the MLP system as a verification device, as opposed to a recognition device, proved to be more feasible especially using the time-adaptive algorithm which could be configured for a FAR of 0.8% so long as an increased FRR of 9% could be tolerated. However, operation at this threshold level may result in high levels of frustration since the resultant FRR means that almost one person will be wrongly rejected for every ten attempts. These levels of FAR and FRR were obtained by setting the decision threshold at 0.5 which represented a low confidence decision.

It must be noted, however, that these performance results were obtained using an effective sensor array of just five sensors instead of the sixteen sensors which were predicted as necessary. Analysis of the discrimination contribution of the feature data for each sensor revealed that 70% of the sensors were either producing non-discriminatory data or were producing similar data and hence were redundant. Recognition performance would be expected to increase significantly if a full complement of sixteen sensors, each contributing different discriminatory information, were available.

Although these rates may not meet current commercial requirements they still represent an impressive starting point into a biometric discipline which is reliant on many new and innovative technologies. Odour could eventually provide the ideal biometric system since if sensors could be developed to respond to minute concentrations of unique human odour compounds then a user could be unaware of assessment as opposed to other biometric systems which currently require either physical contact or a correct orientation such as facial recognition.

Realistically, a biometric access device based upon human odour, using the currently available sensor technology, could only be produced if the technology is integrated with another biometric method in order to provide a higher degree of verification accuracy. This integration of biometric information may be almost undetectable for the access control system user if an appropriate biometric is chosen. For example, if the hand is used as the odour source then a hand profile biometric system could be integrated into the odour sampling system. The user then only perceives one operation being performed: the placement of his/her hand on a suitable dual sampling interface. This hybrid system may also reduce the time taken to sample a user because of increased confidence owing to the combination of two biometrics. This is particularly significant since the odour sampling period is currently approximately twenty two seconds.

The previous comments regarding the difficulties in acceptance of biometric technologies highlight the uphill struggle which faces an entirely new biometric access control method. Electronic noses are currently in their infancy with several companies and educational establishments producing commercial instruments, but at present these are merely used for evaluation purposes and they have not yet proved reliable in most applications. An odour based access control method must not only conquer sceptical attitudes and gain acceptance as an intelligent olfactory machine in industry but must also prove that such a device can reliably function in an access control environment.

7. Conclusion

This work was concerned with the design and analysis of an electronic olfactory device to detect human odour. The device was intended for use in a commercial biometric access control system where a user must be authenticated as a legitimate subject of the access control system using only a unique biological identifier. The following attributes can be deemed as necessary in the construction of a successful biometric device, and conclusions regarding the effectiveness of this work can be drawn from these factors:

- Remote sensing of the biometric feature enabling assessment without the knowledge of the user.
- High speed sensing and computation to reduce inconvenience to the user.
- Ideally operation should be in biometric recognition mode to avoid the use of PINs or ancillary tags.
- Very high levels of reliability to ensure legitimate users are not rejected and intruders are rejected.

A prototype device was designed and constructed which was used to conduct a small field trial. Conclusions regarding the suitability of the device and also conclusions specific to observations made regarding the data will now be discussed.

Measured odour concentrations were not high enough to enable remote sensing of human odour as a result of low human odour levels, environmental effects and low sensors responsivity, but could be implemented in the future if sensor technology progressed sufficiently. Consequently, a static sampling method was employed which utilised the region of initial relatively high concentration found near to the human odour source, the hand. The human odour sample time lasted twenty two seconds, a period which was necessary due to sensor adsorption and desorption time periods. For application in an access control system, it can be concluded that this period was too long to satisfy frustration levels of users, although the ease of use and non-invasive method were positive factors.

Derived sensor features other than the usual maximum divergence increased recognition performance considerably. Primarily, the features of adsorption and desorption gradient contributed extra discriminatory information, which confirmed the predicted effectiveness of these attributes by other researchers. The environmental effects of temperature and humidity were found to affect sensor baseline readings but no correlation with changes in response levels was discovered. From this it was concluded that if environmental levels were more variable, in a commercial application, then environmental considerations would have to be given more importance; these factors had minimal effect in this work.

Analysis of the field trial data revealed that the odour responses exhibited variance with time, an increasing divergence with time was also noted. It can be concluded that this time variance represents a significant problem when considering a commercial biometric system, which uses only data from an initial enrollment period for generating pattern recognition rules. Attempts to find the specific causes of this time variance were inconclusive but they were believed to be a combination of sensor and human causes. It can be concluded that in order to produce a robust system then the effect of time variance must be reduced by refinement of sensor technology, the use of stable human odours and also the adaptation of the pattern recognition system.

The highest obtainable biometric recognition rate was 89% and was produced using an multi-layer perceptron (MLP) neural network and a time adaptive algorithm, which reduced the effect of the previously mentioned time variance of the odour responses. Neural network pattern recognition techniques were shown to be more suitable than statistical methods, for use with the electronic nose used in this work. This superiority was attributed to the effectiveness of neural networks at modeling the non-linearities and distributions within the data. This confirms other work using electronic noses which highlights the effectiveness of neural networks. The majority of the misclassifications experienced were caused by a small number of people, a conclusion in accordance with other work in this field of biometrics.

The best recognition rate is significantly lower than current commercial biometric devices, which typically exhibit rates greater than 99%. It can be concluded that it was unfeasible to use the electronic nose as a commercial biometric recognition device considering the current performance rate. However, use as a verification device was feasible but high user frustration levels would ensue due to high false rejection rates (FRR) of 9%, which were produced at acceptable levels of false acceptance rate (FAR) 0.8%. At this time, most biometric devices operate in verification mode in order to reduce computation times to a minimum and also to improve performance.

A successful biometric system requires discrimination between far greater numbers of people than involved in this work. Performance levels may consequently decrease proportionately and it is this area which begins to highlight the current limitations of the electronic nose for use as a biometric device. Realistically, a biometric access control device based upon human odour, using the currently available sensor technology, could only be produced if the technology were to be integrated with another biometric method in order to provide a higher degree of accuracy. For example, if the hand is used as the odour source then a hand profile biometric system could be integrated into the odour sampling system.

Odour could eventually provide the ideal biometric system since if sensors could be developed to respond to minute concentrations of unique human odour compounds then a user could be unaware of being assessed, unlike other biometric systems which currently require either physical contact or a correct orientation such as facial recognition. Electronic noses are currently in their infancy with several companies and educational establishments producing commercial instruments, but at present these are merely used for evaluation purposes and they have not yet proved reliable in most applications. It therefore follows that an odour based access control method must gain widespread acceptance and must prove that such a device can function with reliability in order to achieve future success.

8. Future Work

8.1 Sensor Performance and Reliability

Conducting polymer sensors are currently the subject of much research and recent work indicates that significant advances are being made so new conducting polymer sensor technologies may be beneficial. The manner in which the sensors are interrogated is also the subject of much research so a more appropriate technique may be applicable. The application of alternative gas sensors, which demonstrate responses to organic odours, could also be investigated.

8.2 Human odour analysis

The identification of genetically unique human odour compounds will be essential in order to produce sensors with greater specificity in the future. This would ensure responses are directly linked to genetically unique compounds and not to naturally varying odour compounds. The identification of these compounds coupled with advancements in sensor specificity could lead to significant performance increases.

The effect of various contamination and poisoning odours must be fully assessed in order to ensure that an odour based biometric system is reliable and robust. This area has been partially addressed in this work to ensure that common perfumes do not contribute to human response levels. The poisoning effect of relevant odour compounds needs also to be addressed. Such odours may permanently or temporarily alter the response characteristics of a sensor.

8.3 Human Odour Sampling

Standards for sampling human odour have not yet been devised. A method should be devised to reduce response variability caused by sampling and also to reduce the influence of environmental effects. The method of sampling may be dependent upon the current sensor technology used. The sampling system may also incorporate filters and liquid traps to reduce the possibility of contamination caused by corrosive volatile odours such as petrol and ammonia.

8.4 Access Control Strategy

Realistically, a biometric access device based upon human odour, using the currently available sensor technology, could only be produced if the technology were to be integrated with another biometric method in order to provide a higher degree of verification accuracy. This integration of biometric information may be almost undetectable for the access control system user if an appropriate biometric is chosen. For example, if the hand is used as the odour source then a hand profile biometric system could be integrated into the odour sampling system. The user would then only perceive one operation being performed, the placement of his/her hand on a suitable dual sampling interface. This hybrid system would also reduce the time taken to sample a user due to the increased confidence owing to the combination of two biometrics.

8.5 Signal Processing

Various aspects of the pattern recognition methods must be addressed in order to ensure that a robust system is produced. The effect of sensor failure must be analysed further to ensure that a small number of sensor failures does not result in complete recognition failure. Novel data rejection should also be investigated further so that behaviour can be predicted when an entirely new data set such as an intruder is introduced. Other refinements which require investigation include effective and fast training algorithms and an efficient method to modify network weights if a user

template requires updating. Ideally a system would measure and classify the odour source simultaneously so that sampling could end as soon enough class membership confidence was obtained; the use of adsorption gradient could provide this dynamic ability.

Field trials of significantly higher numbers of people than used in this study must be performed in order to verify the performance of all aspects of this system. This is especially relevant to such a new type of biometric access control which needs considerable proof of authenticity before it can be accepted as a competitor to other more established biometric methods.

8.6 Time Variance of Odour Responses

The time variance of odour responses has restricted the recognition performance of the electronic nose used throughout this work. This area must be addressed in order to ensure reliable operation of a commercial biometric system based upon odour sensing technology. This problem is complex since many different factors contribute to the time variance of the odour responses and although it is possible that time variance may not be eliminated, steps must be taken to reduce this effect. Firstly, the response variation of the sensors must be investigated to ensure that sensors provide reliable and consistent representations of odours over time. Possible factors influencing sensor response are sensor aging, environmental effects, electronic addressing method, odour interference and odour poisoning. Secondly, the variation of human odour must be investigated to determine the natural variations of the unique odour compounds in question. The specificity of the sensors may have to be refined to eliminate contributions from human odours which exhibit high temporal variations caused by biological and external influences, for example diet, state of health and climate.

9. Appendices

9.1 Sensor Addressing

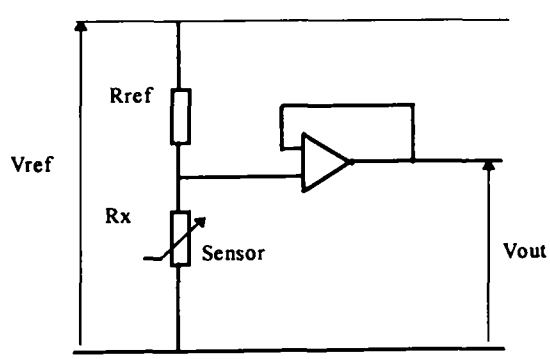
The conducting polymer sensors were interrogated by monitoring its change in resistance by recording the current which flows through the sensor. For a linear resistance sensor the output *Rx* can be expressed in terms of a baseline resistance *Ro*, when the measurand signal is zero, and its fractional change in value *x*, where

$$R_x = R_0(1+x)$$

For negative values of *x*, the ideal range of *x* takes a value between 0 and -1 so that the sensor output covers the full resistance range with the resistance falling to zero when the input signal is at a maximum. In practice, *x* never adopts this ideal state and the exact design of sensor interface depends upon the expected range (and sign) of *x*.

If the resistance change of the sensor, when subjected to a stimulus, is large then a simple potential divider or bridge circuit can be used to process the sensor signal. Figure 10.1 shows a potential divider circuit with a reference resistor, *Rref*, placed in series with the sensor of unknown resistance, *Rx*. *Rref* is chosen to produce a zero output voltage, *Vout*, which allows the data acquisition system to measure either an increase or decrease in resistance of the sensor.

Figure 10.1 Potential Divider Circuit with buffer



The output voltage, *Vout*, from the potential divider, via a unity gain buffer, can be expressed in the following manner :

$$V_{out} = V_{ref}\left(\frac{R_x}{R_x + R_{ref}}\right)$$

To determine the optimum sensor interface it is necessary to analyse the characteristics of the conducting polymer gas sensors. The baseline resistance, *Ro*, of the sensors varies greatly depending upon the particular polymer in question. Baseline resistance of a sensor has been seen to vary from several Ohms to Hundreds of Mega Ohms. A tolerance must be placed on sensor baseline resistance in order to produce a standard electronic interface for any sensor; this tolerance is currently set at 10 KOhm to 80 MOhm.

10. References

1. C.Jennings, 'Biometrics - When the Person is the Key', *Sensor Review*, Vol. 12 Iss. 3, 1992, pp 9-11.

2. R.Terlaga, 'Is your LAN Safe ?', *United States Banker*, Vol. 98 Iss. 7, July 1989, pp 50-52.

3. 'A Necessary Evil Electronic Access Control', *The Biometrics Digest*, March 1997.

4. H.M.Gladney, 'Access Control for Large Collections', *ACM Transactions on Information Systems*, Vol. 15 No. 2, April 1997, pp 154-194.

5. Commercial Information, *Chubb Security Ltd.*, Great Britain, 1994.

6. Commercial Information, *EyeDentify Inc.*, USA,1994.

7. T.Wilson, 'AQ & A Approach', *Security Management*, 1989, pp 23-24.

8. J. Hollingum, 'Automated Finger Print Analysis Offers Fast Verification', *Sensor Review*, Vol. 12 Iss. 3, 1992, pp 12-15.

9. N.Intrator, D.Reisfeld, Y.Yeshurum, 'Face Recognition using a Hybrid Supervised/Unsupervised Neural Network', *Pattern Recognition Letters*, Vol. 17, 1996, pp 67-76.

10. J.R.Parks, 'Biometrics : the people sensors', *Sensor Review*, Vol. 9 Iss. 2, 1989, pp 79-84.

11. M.C.Fairhurst, 'Signature Verification Revisited : Promoting Practical Exploitation of Biometric Technology', *IEE Electronics & Communications Engineering Journal*, Vol. 9 No. 6, Dec 1997, pp 273-280.

12. S.McClelland, 'The Penny Drops for Magnetic Sensors', *Sensor Review*, Vol. 9 Iss. 3, 1989, pp 156-166.

13. C.H.Fander, 'Smart Cards', *Scientific America*, August 1996.

14. M.Marsh (Editor), 'Transponder Background Information', *Transponder News*, Kengray, South Africa.

15. S.G. Davies, 'Touching Big Brother : How Biometric technology will fuse flesh and machine', *Information Technology and People*, Vol. 7. No. 4, 1994.

16. J.Ashbourn, 'Emerging Technology for Security and Control', *Sensor Review*, Vol. 14 No. 4, 1994, pp 3-5.

17. SJB Services, *Biometric Technology Today*, Vol. 2 No. 3, June 1994.

18. M.H.George and R.A.King, 'A Robust Speaker Verification Biometric', *IEEE International Carnahan Conference on Security Technology (29)*, 1995.

19. SJB Services Ltd., *Biometric Technology Today*, Vol. 3 No. 3, June 1995.

20. CBN, 'Biometrics ; Chipping Away Your Rights?', *700 Club Newswatch*, October 1995, http://www.cbn.org/factsheets/biometrics.html.

21. Christian Association Inc., 'Almost Midnight', 1996.

22. R.Sandhu, 'Transformation of Access Rights', *IEEE Transactions on Computers*, 1989, pp 259-268.

23. AfB (Association for Biometrics), 'Glossary of Biometric Terms', 1998.

24. J.D. Daugman, 'Recognising People by their Iris Patterns', *IrisScan Inc. Technology Documentation*, http://www.iriscan.com/basis.htm.

25. D.Ferraiolo and R.Kuhn, 'Role Based Access Control', *Proceedings of the 15th National Computer Security Conference*, Vol. 2, 1992, pp 554-563.

26. R.Sandhu, 'Role Based Access Control', *ACM Transactions on Information Systems*, Vol. 29 No. 2, Feb. 1996, pp 38-47.

27. P Haughton, 'Sensors : Principles and Applications', *Prentice Hall*, 1991, ISBN 0-13-805789-3.

28. Commercial Information, *Mastiff Electronic Systems Ltd.*, Aldershot, Hants., UK, http://www.mastiff.co.uk.

29. Commercial Information, *Identec Ltd.*, Tyne & Wear, England 1995.

30. Commercial Information, *LaserCard Systems Corporation*, Mountain View, CA, USA, 1996.

31. Commercial Information, *Gemplus*, USA, 1996.

32. 'Smart Card Manual', *Xicor Corporation*, USA, 1995.

33. SJB Services, *Biometric Technology Today*, Vol. 2 No. 8, Jan. 1995.

34. Commercial Information, *Iriscan Inc.*, USA, 1994.

35. Q.Jiang, 'Principle Component Analysis and Neural Network based Face Recognition', PhD Thesis, University of Chicago, Department of Computer Science, 1997.

36. Commercial information, *Z.N (Zentrun fur NeuroInformatik) GmBh*, Bochum, Germany, 1995.

37. SJB Services, *Biometric Technology Today*, Vol. 2 No. 1, April 1994.

38. Commercial Information, *East-Shore Technologies Inc.*, Delanson, New York, USA, 1997.

39. Commercial Information, *VeinCheck*, Joe Rice, Nottingham, England, (http://innotts.co.uk/~joerice/), 1994.

40. Commercial Information, *Betac Technology Recognition Systems Inc.*, Alexandria, Virginia, USA, 1998.

41. Commercial Information, *Quintet Inc.*, Cupertino, California, USA, 1996.

42. Commercial Information, 'Biopassword', *NetNanny Inc.*, USA, 1996.

43. B.Gibbons, 'The Intimate Sense of Smell', *National Geographic*, October 1987, pp 324-360.

44. T.C.Pearce, J.W.Gardner, S.Friel, P.N.Bartlett and N.Blair, 'Electronic Nose for Monitoring the Flavour of Beers', *Analyst*, Vol. 118, April 1993, pp 371-377.

45. D.Lancet and U.Pace, 'The Molecular Basis of Odor Recognition', TIBS, Vol. 12, Feb. 1987, pp 63-66.

46. 'Stereochemical Theory of Odor', *Journal of Steroid Biochemistry & Molecular Biology*, Vol. 39, 1991.

47. 'Application Notes and Information about Gas Sensors', *Custom Sensor Solutions Inc.*, Naperville, IL, USA, 1997.

48. H.V.Shurmer and D.J.Whitehouse, 'The Characteristion of Volatile Molecular Substances', *Proceedings from the Royal Society in London*, Vol. A, 1993, pp 313-332.

49. K.Persaud and G.H..Dodd, 'Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose', *Nature*, Iss. 299, 1982, pp 352-355.

50. J.W.Gardner, 'A brief History of Electronic Noses', *Sensors and Actuators B*, Vol. 18-19, 1994, pp 211-220.

51. M.Schleidt and B.Hold, 'Human Odour and Identity', *Olfaction and Endocrine Regulation*, 1982, pp 181-194.

52. I.L.Brisbin, J.R.Austad and S.N.Austad, 'Testing the individual odour theory of canine olfaction', *Animal Behaviour*, Vol. 42, 1991, pp 63-69.

53. J.N.Labows, K.J.McGinley and A.M.Kligman, 'Perspectives on Axillary Odor', *Journal of the Society of Cosmetic Chemistry*, Vol. 34, July 1982, pp 193-202.

54. P.Wallace, 'Brief Communication - Individual Discrimination of Humans by Odor', *Phsiology and Behaviour*, Vol. 19, pp 577-579.

55. R.Ellin, R.Farrand, F.Oberst, C.Crouse, N.Billups, W.Koon, N.Musselman and R.Sidell, 'An apparatus for the detection and quantitation of volatile human effluents', *Journal of Chromatography*, Vol. 100, 1974, pp 137-152.

56. S.D.Sastry, K.T.Buck, J.Janak, M.Dressler and G.Preti, 'Volatiles Emitted by Humans', *Biochemical Applications of Mass Spectrometry*, Wiley, 1980, pp 1085-1133.

57. B.Sommerville and M.Green, 'The sniffing detective', *New Scientist*, Vol. 122 No. 1665, May 1989.

58. B.Sommerville, 'Human Odour Investigation : Progress Report', University of Leeds, Department of Biochemistry, 1989.

59. K.Pope, 'Technology Improves on the Nose As Scientists Try to Mimic Smell', *Staff Reporter of The Wall Street Journal.*

60. P.N.Bartlett and J.W.Gardner, 'Odour Sensors for an Electronic Nose', *Sensors and Sensory Systems for an Electronic Nose* , Kluwar Academic Publishers, 1992, pp 31-51.

61. J.W.Gardner and P.N.Bartlett, 'Potential Applications of Electropolymerised thin organic films in nanotechnology', *Nanotechnology*, Vol. 2, 1992, pp 19-32.

62. P.I.Neaves and J.V.Hatfield, 'A new Generation of Integrated Electronic Noses', *Eurosensors VIII*, 1994, pp 224 - 231.

63. D.Hodgins, 'The Electronic Nose using conducting polymer sensors', *Sensor Review*, Vol. 14 No. 4, 1994, pp 28-31.

64. D.Hodgins, 'The Development of an Electronic Nose for Industrial and Environmental Applications', *Sensors and Actuators B*, Vol. 26-27, 1995, pp 255-258.

65. K.C.Persaud, P.A.Payne, M.E.H.Amrani and G.King, 'Improved Chemical Sensing Characteristics of Conducting Polymers Interrogated at High Frequencies', *IEE Colloquium : Advances in Sensors*, Dec. 1995.

66. P.N.Bartlett and S.K.Ling-Chung, 'Conducting Polymers Part III: Results for four different polymers and five vapours', *Sensors and Actuators*, Vol. 19, 1989, pp 287-292.

67. J.W.Gardner and P.N.Barlett, 'Performance Definition and Standardisation of Electronic Noses', *Proceedings from Eurosensors IX*, Vol. 1, 1995, pp 671-674.

68. H.V.Shurmer, 'Basic Limitations of an Electronic Nose', *Sensors and Actuators*, 1990, pp 48-53.

69. P.Corcoran, 'The effects of Signal Conditioning and Quanisation upon Gas and Odour Sensing System Performance', *Univesity of Derby*, School of Engineering, 1993.

70. 'Development of Conducting Polymer Based Sensors for use in the Detection of Body Odour for Access Control', Final Project Report, Department of Biochemistry and Molecular Biology, Leeds University, Vol. 1, June 1996.

71. M.Schweizer-Beberich, J.Goppert, A.Hierlemann, J.Mitrovis, U.Weimar, W.Rosensteil and W.Gopel, 'Application of Neural Network Systems to the Dynamic Response of Polymer based Sensor Arrays', *Sensors and Actuators B*, Vol. 26-27, 1995, pp 232-236.

72. R.H.Anholt, 'Primary Events in Olfactory Reception', *TIBS*, Vol. 12, Feb. 1987, pp58-60.

73. G.Horner and R.Muller, 'Desired and Achieved Characteristics of Sensor Arrays', *Sensors and Sensory Systems for an Electronic Nose*, NATO ISO Series, ISBN 0-7923-1693-2.

74. J.Hulbert, 'Production of Type 004 Sensors : The Variation in Resistance of Type 004 Sensors with Changes in Humidity', University of Leeds, Department of Biochemistry, 1996.

75. Fukunaga, 'Introduction to Statistical Pattern Recognition', 1982, ISBN 0-12-269850-9.

76. T.Masters, 'Practical Neural Network Recipes in C++', *Academic Press Inc.*, ISBN 0-12-479040-2

77. B.G.Tabachnick, L.S.Fidell and H.Collins, 'Using Multivariate Statistics', 1989, ISBN 0-06-046571-9.

78. J.F.Hair, R.E.Anderson, R.L.Tatham and W.C.Black, 'Multivariate Data Analysis', 1995, ISBN 0-02-349020-9.

79. M.Nadler and E.P.Smith, 'Pattern Recognition Engineering', *John Wiley & Sons*, 1993, ISBN 0471622931.

80. W.S.Meisel, 'Computer Oriented Approaches to Pattern Recognition', *Academic Press*, 1972, ISBN 0-12-488850-X.

81. Duda and Hart, 'Pattern Classification and Scene Analysis', *John Wiley & Sons*, 1973, ISBN 0-471-22361-1.

82. T.A.Skotheim (Editor), 'Handbook of Conducting Polymers', Vols. 1 & 2, Marcel Decker Publishing, New York.

83. G.H.Dunteman, 'Principal Components Analysis', Sage University Paper, 1989.

84. B.S.Everitt and G.Dunn, 'Applied Multivariate Analysis', 1991, ISBN 0-340-54529-1.

85. J.W.Sammon, 'A Non-linear mapping for data structure analysis', *IEEE Transactions on Computers*, Vol. C-18 (5), 1969, pp 401-409.

86. J.Mao and A.Jain, 'Artificial Neural Networks for Feature Extraction and Multivariate Data Projection', *IEEE Transactions on Neural Networks*, Vol. 6 No. 2, 1995, pp 296-317.

87. C.M.Bishop, 'Neural Networks for Pattern Recognition', *Oxford University press*, 1995, ISBN 0-198-53864-2.

88. P.Devijver and J.Kittler, 'Pattern Recognition : A Statistical Approach', Prentice Hall, 1982, ISBN 0136542360.

89. R.Beale and T.Jackson, 'Neural computing an introduction', IOP Publishing Ltd., 1990, ISBN 0852742622.

90. C.G.Looney, 'Pattern recognition using neural networks : theory and algorithms for engineers and scientists', Oxford University Press, 1997, ISBN 0195079205.

91. Y.H.Pao, 'Adaptive Pattern Recognition and Neural Networks', Addison-Wesley Publishing Co., 1989.

92. 'NeuralPro II : Technical Manuals', Neuralware Inc., USA.

93. W.Schiffman, M.Joost and R.Werner, 'Optimisation of the Backpropagation Algorithm for Training Multilayer Perceptrons', University of Koblenz, Institute of Physics, 1994.

94. P.Lisboa, 'Neural Networks', The University of Liverpool, Institute of Advanced Scientific Computation, 1985.

95. 'Local Area Networks. Fortification', *Computer Fraud and Security*, Vol. 11 Iss. 8, 1989, pp 11-14.

96. M.Kochanski, 'How safe is it ?', Vol. 14 Iss. 6, June 1989, pp 257-264.

97. J.D.Gardner, 'Microsensors : Principles and Applications', Wiley, 1994, ISBN 0-471-94136-0.

98. Alder and Roessler, 'Introduction to Probability and Statistics', W H Freeman, 1977.

99. D.Lowe and M.Tipping, 'Feed Forward Neural Networks and Topographic Mappings for Exploratory Data Analysis', *Neural Computing & Applications Journal*, Vol. 4 No. 2, pp 1996, 83-95.

100. J.G.Taylor, 'Using Neural Networks', *Centre for Neural Networks*, Kings College, London.

101. B.V.Dasarathy, 'Decision Fusion Strategies in Multisensor Environments', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21 No. 5, 1991.

102. 'Microchip PIC Embedded Control Handbook',1995, http://www.microchip.com.

103. C.Chong, K.Chang and Y.Bar-Shalom, 'Joint Probabilistic Data Association in Distributed Sensor Networks', *IEEE Transactions Automatic Control*, Vol. 31 Iss. 10, 1986, pp 889-897.

104. J.W.Gardner and P.N.Bartlett, 'Pattern Recognition in Odour Sensing', *Sensors and Sensory Systems for an Electronic Nose*, NATO ISO Series, ISBN 0-7923-1693-2.

105. L.Buck and R.Axel, 'A Novel Multigene Family May Enecode Odorant Receptors: A Molecular Basis for Odor Recognition', *Cell*, Vol. 65, April 1991.

106. M.Horn, 'A new Theory of Absorption for the Quantitive Description of Gas Sensors', *Eurosensors VIII*, 1994, pp 217-219.

107. K.D.Schieraum, U.Weimar and W.Gopel, 'Multicomponent gas Analysis : An Analytical Chemistry Approach Applied to Modified SnO2 Sensors', *Sensors and Actuators*, 1989, pp 71-78.

108. A.Mierzwinksi and Z.Witkiewicz, 'Piezoelectric Detectors Coated with liquid-Crystal materials', *Talanta*, Vol. 34 No. 10, 1987, pp 865-871.

109. J.R.Ullmann, 'Pattern Recognition techniques', *Butterworths*, 1973, ISBN 0-408-70441.

110. M.J.Russel, 'Human Olfactory Communication', *Nature*, Vol. 260, 1976, pp 520-522.

111. M.Holmberg, F.Winquist, I.Lundstrom, J.W. Gardner and E.L. Hines, 'Identification of Paper Quality using a Hybrid Electronic Nose', *Sensors and Actuators B*, pp 246-249.

112. M.Josowicz and P.Topart, 'Studies of the Interactions between Organic Vapours and Organic Semiconductors. Applications to Chemical Sensing', Sensors and Sensory Systems for an Electronic Nose, *NATO ISO Series*, ISBN 0-7923-1693-2.

113. J.J.Cowley and B.W.L.Brooksbank, 'Human Exposure to Putative Pheromones and Changes in Aspects of Social Behavior', *Journal of Steroid, Biochemistry and Molecular Biology*, Vol. 39 No. 4B, 1991, pp 647-659.

114. A.Dravniels, 'Evaluation of Human Body Odors : Methods and Interpretations', *Journal of the Society of Cosmetic Chemistry*, Vol. 26, pp 551-571.

115. A.Coghlan, V.Kiernan and J.Mullis, 'Nowhere to Hide', *New Scientist*, March 1998.

116. D. Stankovic and M.Zlatanovic, 'Versatile Computer Controlled Measuring System for Recording Voltage Current Charcateristics of Various Resistance Sensors', *Eurosensors*, 1993, p 421.

117. A.Hierlemann, U.Weimar, G.Kraus, M.Schweizer-Berberich and W.Gopel, 'Polymer based Sensor Arrays for the Detection of Hazardous Organic Vapours in the Environment', Sensors and Actuators B, Vol. 26-27, 1995, pp 126-134.

118. 'Development of Conducting Polymer Based Sensors for use in the Detection of Body Odour for Access Control', Final Project Report, Department of Biochemistry and Molecular Biology, Leeds University,Vol. 2 Appendices, June 1996.

119. D.Evans, AIPR Paper, Betac TRS (Technology Recognition Systems), Alexandria, Virginia, USA, 1996.

120. B.Erbe and J.Shiner, 'How will biometrics affect privacy rights?', *Scripps Howard News Service*, 1998.

121. P.N.Bartlett, J.Gardner and R.G.Whitaker, 'Electrochemical deposition of Conducting Polymers onto electronic Substrates for Sensor Applications', *Sensors and Actuators A*, Vol. 21-23, 1990, p 911.