

Electronic noses inter-comparison, data fusion and sensor selection in discrimination of standard fruit solutions

P. Boilot^{*}, E.L. Hines, M.A. Gongora, R.S. Folland

*Electrical & Electronics Engineering Division, Intelligent Systems Engineering Laboratory,
School of Engineering, University of Warwick, Coventry CV4 7AL, UK*

Received 12 October 2001; accepted 5 August 2002

Abstract

Intensive research and fast developments in electronic nose (EN) technologies provide the users with a wide spectrum of sensors and systems for their applications. This paper presents some of the results obtained with four different ENs on a series of collaborative tests carried out on six standard fruit samples, pure liquids and mixtures. These experiments, part of the EU ASTEQ concerted action, were designed for inter-comparison of the system's performances. Various feature extraction techniques are considered along with inter-comparison of the individual results obtained with radial basis function (RBF) and probabilistic neural networks (PNN). A low-level data fusion technique is used to merge the various datasets together, considering all extracted parameters in order to increase the amount of information available for classification. We achieve 86.7% correct classification with the fusion system, which outperforms the results obtained with individual ENs. With this fusion array, a problem of dimensionality occurs and it is possible to find an optimal array configuration of reduced dimensionality considering a subset of parameters. We report on various parameter selection methods: principal component analysis (PCA) as a mathematical transformation and two types of genetic algorithms (GAs) optimisation as search methods. Various subsets of parameters are selected and all techniques return improved classification rates, 80% with PCA, 96.7% with *6-integer gene GAs* and 93.3% with *72-binary gene GAs*. In order to overcome cost and technology limitations, optimisation techniques can be used to create application specific arrays selecting the best sensors or the correct parameters.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Electronic nose; Feature extraction; Sensor selection; Genetic algorithms; Probabilistic neural network; Problem of dimensionality

1. Introduction

For the past 20 years, intensive research in the field of electronic nose (EN) technologies has led to improvements in sensor technology [1] and multi-parameter extraction techniques [2]. As a result, the user is provided with a wide spectrum of sensors for their applications and to increase the amount of information for the discrimination of samples using multi-sensor arrays. Many sensors of the same type can be found although they produce only slightly different responses, as they are still non-specific. One of the most promising applications of this approach is in the development of application specific instruments, by selecting the optimum sensors. Research should not only focus on sensor technology development and powerful classification

algorithms, but parameter selection, which is considered to be an important intermediate step as a subset of parameters can often provide better discrimination [3]. Ideally, given a reference database of P odours from N sensors or parameters, it should be possible to identify a subset of n sensors that can produce the best possible discrimination of the p samples of interest [4]. Hence, a systematic or structured method for selecting the best sensors, identifying the optimal array configuration or finding the key parameters would be desired to enhance the overall system performance.

In this paper, we report on the use of four different ENs on a series of collaborative tests carried out on six standard fruit samples, pure liquids and mixtures. These experiments, part of the EU ASTEQ concerted action; were conducted using similar experimental set-ups; and were designed for inter-comparison of the performance of the systems [5]. In Section 2, we present the experimental procedure used together with the individual systems and the parameter extraction techniques considered. Individual EN results are presented in Section 3, extending on data fusion techniques used to

^{*} Corresponding author. Tel.: +44-24-76-528146;

fax: +44-24-76-418922.

E-mail address: p.boilot@warwick.ac.uk (P. Boilot).

URL: <http://www.isel.warwick.ac.uk/PBoilot>

merge the datasets from a mix of sensor technologies. In Section 4, principal component analysis (PCA) is used for preliminary data analysis and we emphasise on ‘the problem of dimensionality’ introduced with the creation of the fusion system. Different methods for parameter selection and dimensionality reduction are considered in Section 5. The use of PCA as an implementation of mathematical transformations, and two types of genetic algorithm (GA) optimisation techniques as search methods will also be discussed in the same section. Results are presented and discussed in Section 6. The conclusions in Section 7, discuss the advantages of the mix of technologies and the design of application specific sensor arrays.

2. Experiment, systems and parameter extraction

The week of experiments that took place at the Flanders Centre for post-Harvest Technology of the Catholic University of Leuven in Belgium from 30 January to 5 February 2000, was the second single-site collaborative instrument evaluation and test meeting held by ASTEQ [5]. There were two main aims for the aroma sensors group, one of the four group involved in the project. The first aim of these tests was to compare the performance of the various EN systems for the assessment of the quality of Cox’s apples in terms of mealiness or damage level. The second aim was to discriminate different standard fruit solutions, pure liquids and mixtures, constituting a more basic metrology approach.

2.1. Experimental procedure

For this last series of tests, a lot of attention was given to gathering instruments in the same place and the use of the same type of prepared samples as well as devising a common protocol for headspace generation. These considerations were devised to help us assess repeatability, reproducibility and sensitivity of the EN systems. They were implemented as rigorously as possible to be able to focus on the inter-comparison of the results. For these experiments, six artificial solutions were considered:

- real apple juice;
- pure artificial: pear with ethanol, pear without ethanol and peach;
- mixture: pear 40%/peach 60% and pear 60%/peach 40%.

A fixed volume of liquid, typically 100 ml was introduced into a closed jar filled with control air and the set-up was left to stand undisturbed for 20 min for headspace generation and stabilisation. Control air was also used for the baseline by most of the systems. The different solutions were presented at random to the ENs, each system using different samples every time. The number of samples collected and the number of repetitions for each liquid differed from one EN to the other. Temperature and humidity were

continuously monitored during the tests, and external conditions creating background disturbances were assumed to be affecting all systems in the same way.

2.2. EN systems

From this series of collaborative tests, we will consider a group of four EN systems using different technologies but which employed the same experimental protocol for headspace generation. The data files produced by the operators were gathered and various signal-processing techniques were used to extract the key features or parameters from the datasets. The four EN systems considered are:

- **INRA**: an electronic olfactometer based on a virtual sensor array (INRA Dijon) operated by Mr. Patrick Mielle. The datasets available represented the entire spectral response, including the transient signal for the initial base line, and five zones corresponding to different modulated operating temperatures for the single SnO₂ metal oxide sensor. For each five zones, logarithmic in shape, we extracted the static responses from the mean of the signal (5^* *Mean Df psens*) and also the maximum of the first derivative of the signal in order to explore the dynamic information (5^* *Max Derv psens*). A total of 16 samples were available and 10 parameters were considered.
- **Roma**: a prototype EN of the LibraNose series (University of Rome ‘Tor Vegata’ and Technochip) based on seven ‘thickness shear mode resonators’ (TSMR) operated by Dr. Corrado Di Natale. The datasets available represented the time responses over 8 min of the seven sensors of standard exponentially decreasing shape. As an extracted sensor response the mean of the deflection per sensor was extracted (7^* *Mean Df sens*), as well as the maximum of the first derivative of the sensor signals (7^* *Min Derv sens*). A total of 21 samples were available and 14 features were considered.
- **UPM**: another version of the LibraNose (University of Rome ‘Tor Vegata’ and Technochip) based on eight quartz crystal microbalance sensors (QMBS) operated by Ms. Eva Correa. For this data, already pre-processed responses were available consisting of the mean deflection (8^* *Df sens*) and maximum deflection (8^* *Dfmax sens*) for each sensor. A total of 20 samples were available and 16 features were considered.
- **Warwick**: a prototype of the Cyrano Sciences’ EN, datalogger (Cyrano Sciences Inc.) based on 32 swollen conductive polymer (CP) sensors operated by Mr. Pascal Boilot. For this data, one feature per sensor was extracted using a fractional difference model pre-processing algorithm; $((R - R_0)/R_0)$, using the sample reading (R) and the baseline (R_0) (32^* *FD sens*). A total of 50 samples were available and 32 features were considered.

3. Individual results and data fusion

Radial basis function (RBF) networks and probabilistic neural networks (PNN) were used to perform pattern recognition [6,7]. The relatively small number of vectors per dataset was a problem and it was decided to use these two types of networks as they can help to overcome this limitation, to a certain extent compared to multi-layer perceptron (MLP). When implementing these two classifiers using MATLAB[®], the *spread constant* for the radial basis layer had to be set and multiple values investigated, but only the value producing the best results is reported. Each dataset was normalised using sensor normalisation given in Eq. (1), by considering sensor i and sample j , in order to set the range of each dimension to [0, 1].

$$k_{ij} = \frac{x_{ij} - x_{i,\min}}{x_{i,\max} - x_{i,\min}} \quad (1)$$

Due to the scarcity of data points, validation was facilitated by the leaving one out algorithm in which all patterns are used to train the network except one left out for test. This process is then repeated for all vectors in the dataset. Results for each individual system using the original dataset formed after feature extraction and normalisation are presented in Table 1.

3.1. Low-level data fusion

In order to perform inter-comparison of the EN systems in discrimination of the various fruit solutions, we felt that it was most appropriate to use datasets of the same size, containing the same number of patterns per class. With the original datasets, the number of patterns per set is different from one system to the other, and also there are different numbers of patterns per class for one system. New datasets were formed from the original non-normalised ones, they have an equal number of patterns per class, arbitrarily set to 5, using data reduction and generation techniques. To generate more vectors from *INRA*, *Roma* and *UPM* datasets, we used the original data and additional vectors generated using the mean per class and the mean $\pm 1/2$ standard deviation. For *Warwick* dataset, we employed data reduction by considering the mean per class or the mean of repetitions. After sensor normalisation, RBF and PNN were used as classifiers; the results are shown in Table 1.

For each one of the four ENs, the new datasets formed contained the same number of patterns (30), and they also had the same number of patterns per class (5). A low-level data fusion technique was used to merge the various new datasets together, considering all extracted parameters in order to increase the amount of information available for classification [8]. At low-level fusion, data matrices with the same number of vectors are combined together into a single matrix having a number of rows equal to the number of patterns (30) and a number of dimensions equal to the total number of parameters (72). The 72 dimensions, referred to as *Param1* to *Param72*, were created by assembling individual matrices. We considered first the 10 features from *INRA*, then the 14 ones from *Roma*, the 16 parameters from *UPM* and the 32 sensor responses from *Warwick*. RBF and PNN were then used on the fusion system to evaluate the classification accuracy; the results are shown in Table 1.

3.2. Results

With the original datasets, most of the classification accuracies were below 50%, which indicates that the ENs did not manage to separate the various samples of interest. We achieved 75% correct classification using RBF with the *INRA* data by extracting parameters that best represented the information embedded within the responses. However, for *UPM* and *Roma*, we only achieved 35 and 38%, respectively with RBF and PNN, which may indicate that the algorithms chosen for feature extraction are not the most suitable to best represent the information. In general with the new datasets, classification accuracies have been improved as the data generation techniques used tended to add patterns to a class reducing its overall variance. For example, we achieved 66.7% correct classification with RBF on *UPM* data and up to 80% with RBF on *Warwick* data. Results were lower for *INRA* as there was no sample for pear with ethanol due to the system responses going off the scale. In the new dataset they were replaced by the same patterns used for pear without ethanol, which may explain the confusion in identification. We found that PNN trained with a *spread constant* set to 0.25 always produced the best classification results, this value will be used later when we attempt to select sensors. We achieved 86.7% correct classification with the fusion system, which outperformed the results obtained with individual ENs.

Table 1
RBF and PNN classification results using the original, the new and the fusion datasets

EN system	Original dataset				New dataset				Fusion All
	<i>INRA</i>	<i>UPM</i>	<i>Roma</i>	<i>Warwick</i>	<i>INRA</i>	<i>UPM</i>	<i>Roma</i>	<i>Warwick</i>	
No. of patterns	16	20	21	50	30	30	30	30	30
No. of features	10	16	14	32	10	16	14	32	72
RBF (%) (spread constant)	75 (2)	35 (1.2)	19 (0.1)	34 (1)	56.7 (1.2)	66.7 (2)	40 (10)	80 (5)	86.7 (10)
PNN (%) (spread constant)	62.5 (0.8)	25 (0.5)	38.1 (1.2)	56 (0.25)	53.3 (0.25)	53.3 (0.25)	33.3 (0.25)	43.3 (0.25)	80 (0.25)

4. Problem of dimensionality

The creation of the fusion system allows us to investigate the potential for utilising a mix of sensor technologies in the classification of odours. *Roma* and *UPM* ENs are based on the same sensor technology however, sensors of the same type may produce only slightly different responses and it will be interesting to identify irrelevant or redundant sensors. It is possible to extend this idea to the entire fusion system for which many parameters may be highly correlated.

4.1. Problem of dimensionality

With the fusion system, it was possible to merge a wide spectrum of sensors or extracted features, but it generated a selection problem for this specific application. We were able to create a larger and more complex system, of which the level of performance is better than the individual systems, but by increasing the array size we have introduced redundancy. Therefore, a method for selecting a subset of parameters in order to optimise the performance of the fusion system is desired. In most cases a simpler application of a subset of sensors can provide better recognition rate than the whole set of parameters because most classifiers will not produce optimal results if irrelevant or redundant sensors are used [3]. In an optimisation process, a measure of the correct classification rate can be used to maximise the performance of the fusion array, whilst minimising the number of parameters selected [4]. Reducing the number of parameters in order to maximise the performance is achieved at certain cost, and parameters should be selected carefully as shown in the *configuration performance plot* in Fig. 1. Too few,

inappropriate or poor configuration of parameters can lead to an inadequate array performance. Conversely, increasing the array size and the number of parameters may not be effective if the parameters selected are redundant or do not help to improve discriminatory power.

4.2. PCA preliminary analysis

A PCA was performed on the 72 parameters, as a dimensionality reduction technique, in order to investigate how the patterns clustered in the representation of the ‘fusion multi-sensor space’. In Fig. 2, a plot of the first two principal components, which account for 82.67% of the variance, shows there is no clear discrimination between the various clusters representing the fruit solutions. In the pattern classification space, the high level of similarity between the fruit samples causes clusters to overlap, which indicates the difficulty in separating them and the relatively low level of separation achieved by using the extracted parameters. Clearly, with this problem for which 72 parameters are available, an optimisation process is required to identify a configuration with fewer parameters that will give equivalent or better levels of performance. Reducing the number of sensors is desirable in order to optimise the system cost, speed and reliability and in order to select the best sensors suitable for this application.

5. Parameter selection

The optimisation processes we used focus on method of dimensionality reduction and feature selection for which a subset of parameters of the original space is selected. In this

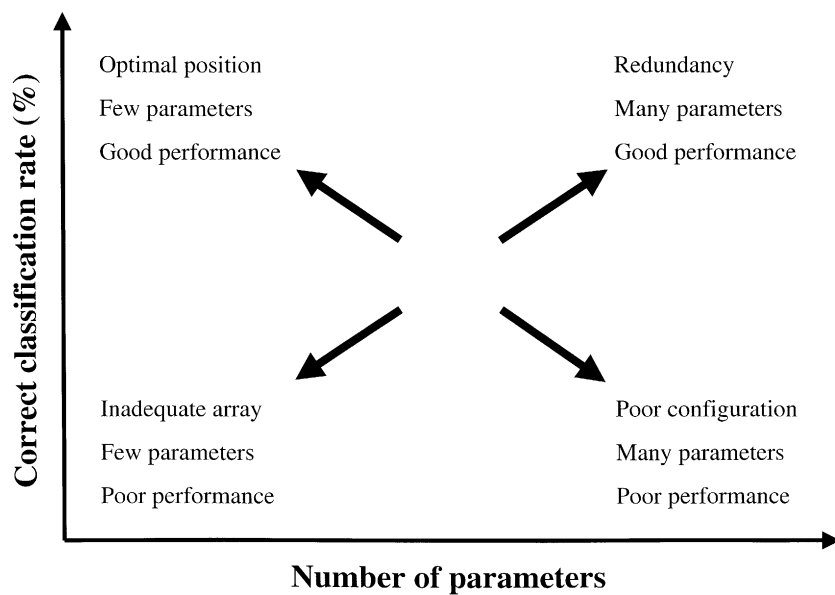


Fig. 1. Configuration performance plot for sensor reduction. (Reprinted from Corcoran, Sens. Actuators A, 76, 57–66, © 1999 Elsevier Science, with permission.)

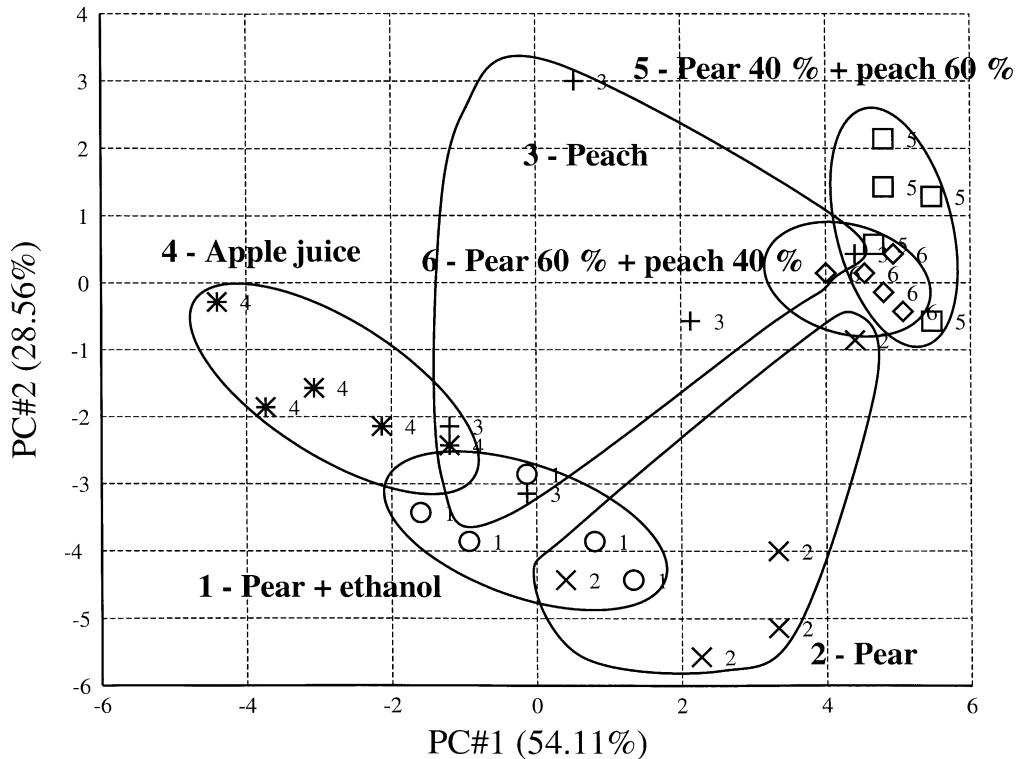


Fig. 2. PCA score plot, clustering representation for fusion system.

section, we report on the use of various parameter selection methods that will be applied to the optimisation of the results in the discrimination of standard fruit solutions from a fusion system. We report on the use of PCA as an implementation of a mathematical transformation based on linear projection and two types of GA optimisation techniques as search-based methods.

5.1. PCA as a projection method

PCA is based on a linear project of multidimensional data onto different coordinates based on maximum variance and minimum correlation [9]. As a result, less significant components can be eliminated, reducing the data representation to only those responsible for the most significant contribution. PCA uses eigenvectors and eigenvalues to define the subspace orthogonal base from the data covariance, preserving as much as possible, the variance presented in the original N -dimensional data into the new subspace. PCA is often used as a dimension reduction technique to map an N -dimensional sensor space onto a two-dimensional feature space, giving an appropriate representation of a dataset. Through PCA, information on the coordinates of the data in the new coordinates can be obtained, from which a *score plot* can be drawn to best study the clustering of data as shown in Fig. 2. Also, the contribution of each sensor to each component can be extracted, from which a *loading plot* can be drawn to give information about the mutual correlation of sensors as shown in Fig. 3.

Using the first two principal components, the distribution of the features from the four EN systems follows an interesting pattern; along PC#2, top for *INRA* features and bottom for *Roma* parameters, and along PC#1, left for *Warwick* sensors and right for *UPM* ones. In order to select a subset of parameters, features with minimum loadings are removed and only parameters that contribute the most to principal components will be considered. This selection technique allows us to eliminate the most orthogonal sensors, selecting the two extreme parameters on one principal component to best represent the maximum variance captured in that direction. Using the first four principal components, 95.49% of the variance of the data is captured, using this technique it is possible to select a subset of eight parameters, as shown in Table 2. The parameters come from all four ENs although the *INRA* system seems to present the best features with four parameters selected out of eight.

5.2. GAs for sensor selection

In principle an exhaustive search for the best result can be conducted, but for this particular 72-parameters application, a total of 2^{72} combinations need to be evaluated. By reducing the problem to only eight parameters would still leave 4.8×10^{14} ($72 \times 71 \times \dots \times 65$) combinations to be considered. Different search algorithms are available, which classical sequential forward and classical sequential backward selections being two examples [10], but they will only explore a small fraction of the whole set of configurations.

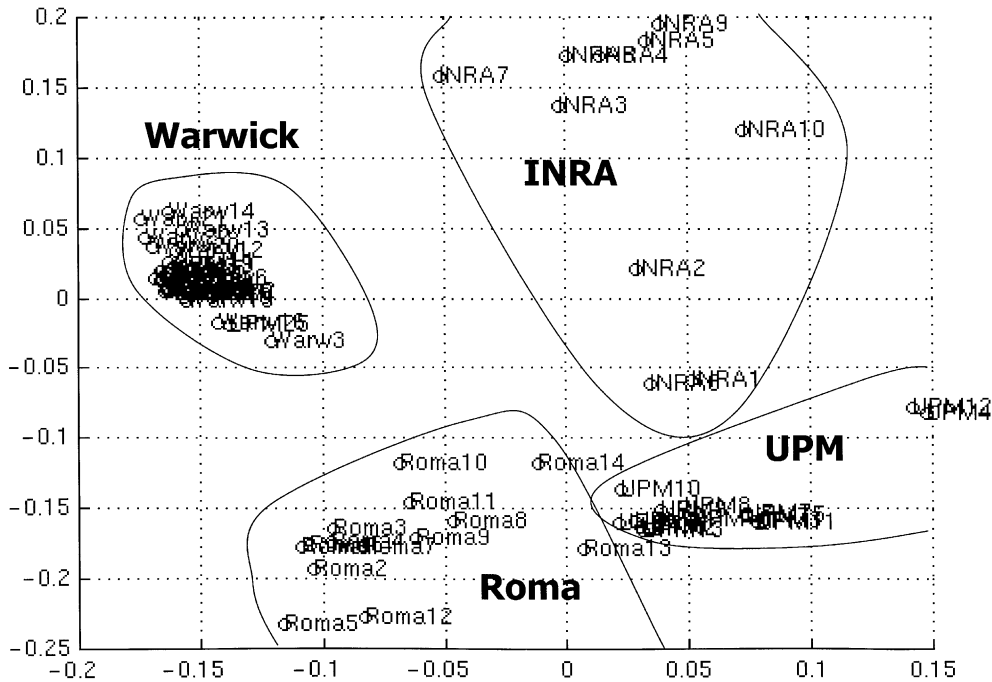


Fig. 3. PCA loading plot, ENs features correlations for fusion system.

However, we will explore the use GA-based search methods using the maximisation or minimisation of an objective function to guide the feature selection, as they have much to offer [11]. The GA is an optimisation technique that will arrive at an optimal solution to a particular problem using the survival of the fittest principle as presented in Darwin’s theory of evolution.

In most cases, when GAs are applied to feature selection problems, the chromosomes representing the entire set of features can easily be codified as a binary string where 1 means that the feature is present and 0 means that the feature is absent [3,4]. The selected parameters for array configuration use a genetic representation for which each parameter is equivalent to a gene. For this particular application,

this representation was adopted with the chromosomes representing the selected parameters being 72-genes long (the results will be referred to as *72-binary genes GA*). In most applications this representation is the most suitable, but for this particular selection problem genes formed might be too complex and too long. Therefore, we decided to apply a new coding technique for the creation of the population of chromosomes. For the genetic representation, we used *X-genes* long chromosomes with integer values from 1 to 72, representing the selected parameters; one of the conditions was that there will be no repetition of one value in the same chromosome (referred to as *check for integrity*). For this particular GA selection, various values for *X*, number of parameters selected, were investigated and the results will be referred to as *8-integer genes GA* or *6-integer genes GA*.

The GAs used follow the procedure described in Fig. 4, so that the best configuration achieving the best recognition rate is found. The initial population was generated at random and the classification rates found at this stage were used to produce the results for randomly-generated subsets. There are no rules in defining the size of the initial population and for *72-binary gene GAs* a population of 15 chromosomes was considered. For optimisation in *X-integer gene GAs*, it was found that the initial population should contain at least twice as many genes as the total number of parameters: 18 chromosomes for *8-integer gene GAs* and 25 chromosomes for *6-integer gene GAs*. The *check for integrity* is only performed for *X-integer gene GAs*. With every generation, the recognition rate is evaluated using PNN for every individual within the population. We used the survival of the fittest (elitist model) to ensure the best chromosome will

Table 2
PCA selected parameters

PC#1 (54.11%)				
Max	0.1473	Param28	UPM	Df sens4
Min	-0.1742	Param71	Warwick	FD sens31
PC#2 (28.56%)				
Max	0.1945	Param9	INRA	Max Derv psens4
Min	-0.2334	Param15	Roma	Mean Df sens5
PC#3 (7.82%)				
Max	0.3271	Param6	INRA	Max Derv psens1
Min	-0.1828	Param17	Roma	Mean Df sens7
PC#4 (5.01%)				
Max	0.4067	Param8	INRA	Max Derv psens3
Min	-0.1828	Param1	INRA	Mean Df psens1

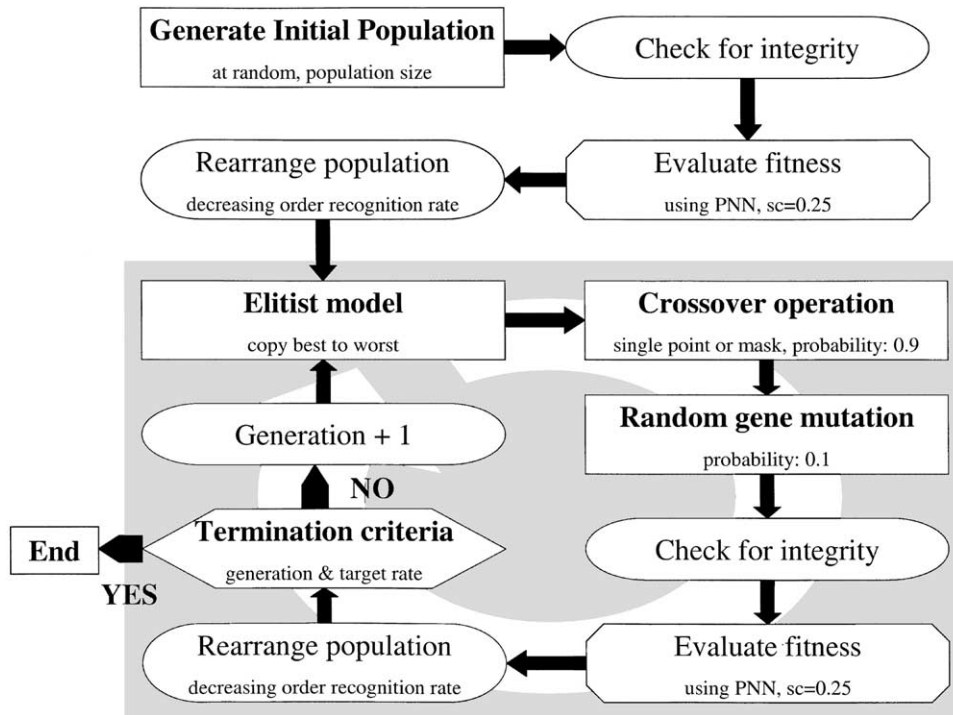


Fig. 4. Schematic depicting the process of GA parameter selection.

survive the next generation and will not be affected by genetic operators. At the rearrangement stage, chromosomes are sorted in ascending order of recognition rate, with a copy of the fittest replacing the worst before genetic operators are applied. For *X-integer gene GAs*, single crossover points are applied in each population; two parents are selected at random and genes are crossed over at a random point to constitute the offspring chromosomes of the new generations. For *72-binary gene GAs*, a mask is created to define which genes should be swapped over. The number of crossovers to apply per generation is set so that the probability of one chromosome to be affected is about 0.9. Random point gene mutations are then carried out on the chromosomes, randomly selecting a gene from the entire population and changing its value. The number of mutations to apply per generation is set so that the probability of one gene to be affected is about 0.1. After applying the genetic operators, the *check for integrity* is performed, and the fitness of each new chromosome is evaluated using the recognition rate from the PNN results. Two optimisation criteria that will terminate the GA optimisation process were selected; the maximum number of generations was set to 20 and the target recognition rate was set to over 95%.

6. Results

All implementations of PCA and GA parameter selection were run under MATLAB[®]. The PNN and PCA used are built-in programs of this package but a customised program

was developed for GAs. PNN, with a spread constant of 0.25 and leaving one out for validation, was used as the classification paradigm to measure the level of performance achieved with a selected subset of parameters. PNN is used because it was shown earlier to give the best results for individual systems and handles relatively well with small datasets. It is believed that other classifiers will perform better in this situation but it was beyond the scope of this investigation to explore this here. For PCA, we considered a subset of eight parameters from the first four principal components and a subset of six parameters using only the first three components. For GA, *X-integer genes* and *72-binary genes*, the results are presented in Table 3, representing the average after 10 different runs for the fittest chromosomes and for all chromosomes in the last population. From the initial population generated at random, results for randomly generated datasets can be extracted.

For the full parameter set we achieved 80% correct classification with PNN, using a subset of eight parameters selected considering the loadings from PCA, we achieved equal levels of system performance whilst reducing array dimensionality by 89%. For *72-binary gene GAs*, further examination of the complete set of results shows that the GA search method consistently identified subsets containing between 20 and 30 parameters. These subsets have optimal classification rates of around 94.3% (96.7% at best) compared with randomly generated data sets, which typically have classification rates around 75.2% (with one exception at 93.3%). With this GA parameter selection technique, the array dimensionality is reduced by 60–70%, identifying

Table 3
Parameter selection results using PNN as classifier

PCA: 4 PCs; 8 parameters; 80%			PCA: 3 PCs; 6 parameters; 70%			
			GA			
8-integer	Average last population	75.7%	Average bests	94%	Best of all	96.7%
6-integer	Average last population	72.1%	Average bests	95%	Best of all	96.7%
72-binary	Average last population	85.3%	Average bests	94.3%	Best of all	96.7%
			Random			
8-integer	Average initial population	66.5%	Average bests	84.7%	Best of all	90%
6-integer	Average initial population	64%	Average bests	86%	Best of all	90%
72-binary	Average initial population	75.2%	Average bests	86.7	Best of all	93.3%

some of the redundant parameters, but still too many parameters are considered and this technique shows its limit for this dimensionally-large fusion array.

The *X-integer gene GAs* were suggested for a faster and a more efficient approach to dimensionality reduction. With *8-integer gene GAs*, the array dimensionality is reduced by 89%. The subsets of parameters found for the fittest chromosome at the end of the optimisation process achieved on average being 94% correct classification (96.7% at best). It is clear that the same level of performance cannot be achieved through random selection of a similar number of parameters, as the classification rates are much lower with only 66.5%. This idea was extended to an even fewer parameters and with *6-integer gene GAs*. The search method could consistently identify subsets of parameters that will achieve high levels of performance: 95% on average, 96.7% at best. Selecting other values for *X* does not improve the system performance any further. No further statistical analyses of selected subsets were conducted to find which parameters were selected the most.

7. Conclusions

For individual ENs, no real generic conclusions can be drawn concerning inter-comparison due to the relatively small datasets available not being very representative of the problem; and the poor selectivity of the feature extraction techniques used to represent the information. A low-level data fusion technique was used to merge the various datasets together, considering all extracted parameters, in order to increase the amount of information available for classification of the six standard liquid samples. Results obtained with the fusion system outperformed the classification results of the ENs taken individually. The use of optimisation techniques for dimensionality reduction and parameter selection was considered in order to find a subset of parameters that will improve the levels of performance. Parameter selection is required in order to develop an application specific system by eliminating redundant or irrelevant features, therefore optimising the array configuration. Using PCA, a subset was selected by considering the

parameters with maximum and minimum loadings for each one of the first four principal components. The PCA subset achieved the same level of system performance as the entire fusion array, namely 80%, but it is understood that the validity of this technique is dependent upon the dataset. GA implementations were considered for parameter selection, using genetic coding of the parameter index in a binary string or as an integer value. With *72-binary gene GAs*, dimensionality is reduced by 60–70%, and we achieved classification rates of around 94.3%; nevertheless the array dimensionality can be further reduced by 89% using *X-integer genes GAs*. These optimisation techniques achieved more than 94% correct classification using subsets of only eight or six parameters. These results were compared with subsets of parameters randomly generated, which typically achieved 65–75% correct classification. *X-integer gene GAs* are believed to be the best, and the fastest, parameter selection paradigm investigated and we believe it will prove to be reliable and beneficial in defining the optimal array configuration. This search method was successfully employed to investigate the potential of utilising a mix for sensor technologies and the selection of an optimal array configuration for an application specific problem. Using sensors from an existing database, the best subset for a given application can be found in this way, reducing the cost associated with new sensor developments, or using redundant sensors in an array.

Acknowledgements

Pascal Boilot gratefully acknowledges financial support (award number 99310943) from EPSRC during his stay as a Ph.D. student at the University of Warwick. We also like to express thanks to the ASTEQ collaborating colleagues, D. Rutledge (ASTEQ co-ordinator), P. Grenier (aroma co-leader), P. Mielle (INRA), E. Correa (UPM), C. Di Natale (Roma) and others without whom this work would not have been possible. The Commission of the European Communities is gratefully acknowledged for financial support (Concerted Action “ASTEQ—Artificial Sensing Techniques for Evaluating Quality” FAIR3 CT97-3516). Contents of this

publication are the sole responsibility of the authors and in no way represent the views of the Commission or its services.

References

- [1] J.W. Gardner, P.N. Bartlett, *Electronic Noses: Principles and Applications*, Oxford University Press, New York, 1999.
- [2] E.L. Hines, E. Llobet, J.W. Gardner, Electronic noses: a review of signal processing techniques, *IEE Proc. Circuits Dev. Systems* 146 (1999) 297–310.
- [3] S. Pardo, S. Marco, C. Calaza, A. Ortega, A. Perera, T. Sundic, J. Samitier, Methods for sensor selection in pattern recognition, in: J.W. Gardner, K.C. Persaud (Eds.), *Electronic Noses and Olfaction*, IoP Publishing, Bristol, 2000, pp. 83–88.
- [4] P. Corcoran, J. Anglesea, M. Elshaw, The application of genetic algorithms to sensor parameter selection for multisensor array configuration, *Sens. Actuators A* 76 (1999) 57–66.
- [5] ASTEQ—Artificial Sensing Techniques for Evaluating Quality, Concerted Action No. FAIR5-CT97-3516, funded by the Commission of European Communities, (online) <http://www.inapg.inra.fr/ens_rech/siab/asteq/index.htm>.
- [6] S. Chen, C.F.N. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, *IEEE Trans. Neural Networks* 2 (1991) 302–309.
- [7] E.L. Hines, P. Boilot, J.W. Gardner, M.A. Gongora, E. Llobet, Pattern analysis for electronic noses, in: T.C. Pearce, S.S. Schiffman, H.T. Nagle, J.W. Gardner (Eds.), *Handbook of Machine Olfaction: Electronic Nose Technology*, Wiley, Weinheim, 2002 (Chapter 8).
- [8] C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, L. Llova, A. Rudnitskaya, Y. Vlasov, Electronic nose and electronic tongue integration for improved classification of clinical and food samples, *Sens. Actuators B* 64 (2000) 15–21.
- [9] J.W. Gardner, Detection of vapours and odours from a multi-sensor array using pattern recognition. Part 1. Principal components and cluster analysis, *Sens. Actuators B* 4 (1991) 108–116.
- [10] S. Stearns, On selecting features for pattern classifiers, in: *Proceedings of the 3rd International Conference on Pattern Recognition*, Coronado, USA, 1976, pp. 71–75.
- [11] B.G. Kermani, S.S. Schiffman, H.T. Nagle, A novel method for reducing the dimensionality in a sensor array, *IEEE Trans. Instrument. Measurement* 47 (1998) 728–741.