Neural network based electronic nose for classification of tea aroma

S Borah · E L Hines · M S Leeson · D D Iliescu · M Bhuyan · J W Gardner

Abstract This paper describes an investigation into the performance of a Neural Network (NN) based Electronic Nose (EN) system, which can discriminate the aroma of different tea grades. The EN system comprising of an array of four tin-oxide gas sensors was used to sniff thirteen randomly selected tea grades, which were exemplars of eight categories in terms of aroma profiles. The mean and peak of the transient signals generated by the gas sensors, as a result of aroma sniffing, were treated as the feature vectors for the analysis. Principal Component Analysis (PCA) was used to visualise the different categories of aroma profiles. In addition, K-means and Kohonen's Self Organising Map (SOM) cluster analysis indicated there were eight clusters in the dataset. Data classification was performed using supervised NN classifiers; namely the Multi-Layer Perceptron (MLP) network, Radial Basis Function (RBF) network, and Constructive Probabilistic Neural Network (CPNN) were used for aroma classification. The results were that the three NNs performed as follows: 90.77, 92.31, and 93.85%, respectively in terms of classification accuracy. Hence the performance of the proposed method of aroma analysis demonstrates that it is possible to use NN based EN to assist with the tea quality monitoring procedure during the tea grading process. In addition the results indicate the possibility for standardization of the tea aroma in numeric terms.

Keywords Tea aroma _ Electronic nose _ Neural network

Introduction

The characteristic fragrance of tea, the 'aroma' of either infused leaf or the tea liquor, is one of the most significant parameters for evaluation of its quality [1–3]. The aroma indicates that the tea comprises at least one of a certain number of odours that are desirable and are highly valued in 'good quality tea'. This odour is also known by the 'tea tasters' as the 'nose' or 'fragrance' and is one of the three attributes of the tea flavour relating to the smell of the tea, and is utilized for the assessment of tea quality in the tea industry. The other two flavour attributes are senses of taste and astringency, which are perceived within the taster's mouth. The source of the tea aroma is attributed to various volatile organic compounds (VOCs) present in tea. These VOCs are formed as a result of significant lipid degradations during the course of the various tea processing stages [4]. For example, the majority of the aroma compounds are formed from amino acids, fatty acids, carotenoids and glycosides [5]. On the other hand, the variation of aroma in different teas is attributed to the variation of these VOCs, and their ratios, as present in tea. There are many reasons for such variations in aroma profile of which one of the main reasons is due to tea being made from different varieties of 'tea clones' [6]. Moreover, the processing techniques, the pruning and the pruning time also have a major impact on the variations in VOCs; thus affecting the aroma profiles [7–9]. During quality evaluation, experienced tea tasters, supported by certain standard flavor terminology, sniff the tea aroma using organoleptic methods and thus play a key role in the assessment of the quality of the different grades of tea. For example, aroma terms 'flat' and 'plain' describe poor tea quality, where flat tea is caused either by damp storage conditions or the relative age of the tea. 'Plain' tea is due to the fact that the tea was produced in a less favorable season of the year, such as during the monsoon the period. Naturally for commercial reasons the most profitable teas are the most important ones; these being the ones which are perceived by the consumer to be the most desirable.

There are usually two distinct aspects for aroma profiling as adopted by the tea tasters. One is the strength of the tea aroma as determined using a scale of 1–10. The other is concerned with categorizing teas in terms of the variability of the tea aroma. Some terms commonly used to describe the aroma include: fresh floral, sweet floral, citrus, sweet fruity, fresh green, sweet, resinous, roasted, dimethyl sulfide- like, green, burned, acidic, fermented, oily, earthy and moldy. The valuation and pricing of the final tea is determined by accounting for the other flavour attributes in the context of these two aspects of aroma profiling. As in other food industries, the tea producers seek to modernize their quality evaluation techniques to satisfy a market driven by customer demands for products with greater consistency, better flavour and product differentiation. As a consequence, the tea producers are interested in the possibilities for efficient on-line evaluation of the flavor profiles. In this context, although the organoleptic method is in use in the tea industry it does not offer the possibility for real time, on-line monitoring of the 'aroma analysing processes'. Moreover, a numerical descriptor or score for the aroma terms used by the tea tasters is not currently available. Therefore, the aroma profile must be standardized at the grading stage in order to facilitate efficient decision making in respect of tea quality. Statistical studies show that there are more than 10 VOCs present in tea, which are responsible for its aroma [6]. Hence this potentially makes the 'aroma analyzing phenomenon' a complex problem, in terms of having to deal with the overall aroma resulting from the combination of VOCs. The phenomenon is further complicated for tea tasters by the fact that the tea aroma is unstable and changing with time as the VOC composition changes during the various tea processing stages. Additionally, some disadvantages of organoleptic techniques for aroma profiling include: individual variability, adaptation (becoming less sensitive due to prolonged exposure), fatigue, infection, mental state etc [10]. These disadvantages of the traditional methods and the need for aroma standardization for on-line quality evaluation process make it worthwhile to explore the potential application of electronic nose (EN) for 'aroma' profiling in the tea industry.

The EN system has emerged over recent years into the limelight due to its outstanding performance in odour detection and classification of various food products [11]. It is a very popular choice for researchers to use in various food processing industries for quality monitoring owing to its unique response to different odours. The EN typically comprises of an integrated chemical sensor array, together with interfacing electronic circuitry and a pattern recognition unit [12, 13]. In the conventional EN setting, researchers typically gather data and adopt statistical methods to analyse them. More recently, novel Intelligent System Engineering (ISE) techniques such as NNs, Fuzzy Logic (FL) and Genetic Algorithms (GAs) have been applied to help optimise the effectiveness of the data processing, classification, and other outcomes [14] by, for example, extracting key features and relevant information in order to draw the best conclusions. The aim of the data processing is to optimally facilitate the realisation of any pertinent information/knowledge which may be forthcoming as a result of the data gathering. Finally; such an approach contributes new knowledge which can help to formulate more universally acceptable aroma terminology. This paper contains the following main sections: Materials and methods, Data analysis, Results and discussions.

Materials and methods

Sample preparation

Thirteen different tea grades were randomly selected for the initial set of experiments. The grades selected were based on organoleptic decisions made by tea tasters, using their records. The tea tasters' judgement of aroma resulted in some samples being assigned the same score even though they were of different grades from different tea gardens; and vice versa. The collected samples were then divided into seven different categories as a result of organoleptic testing by tea tasters, see Fig. 1. The quality of the tea grades are categorized from highest quality to lowest quality, from top to bottom as shown in the figure.

The main procedure used by the tea tasters for sniffing tea aroma is to add hot water to a specific amount of tea sample. This is in fact the conventional way of preparing 'tea liquor' most of the time. While adding hot water, VOCs escape from the tea liquor and this is the preferred way in which the tea taster makes the necessary assessment of the aroma. By analogy with this method of 'sniffing', the same procedure is followed for 'EN based sniffing'. Test samples were prepared by picking 1.5 g from each of the selected tea grades. The tea liquors were produced using each sample by adding 200 ml of boiled distilled water.

Experimental setup

Figure 2 shows the experimental setup of the EN system for tea aroma sniffing. Our EN uses four tin oxide sensors; namely TGS 880, TGS 826, TGS 825 and TGS 822 (Figaro Engineering Inc.). These sensors were selected because they have been found to respond well to the different odours produced by the tea, Table 1 lists the sensors along with their main characteristics. The EN in general terms consists of three functional components that operate serially on an odorant sample. They are sample handler, an array of gas sensors, and a data analysis sub-system. The sample handler is a necessary arrangement to transport the odorant molecules of a sample from the sample collection device to facilitate contact with the sensor headspace. It consists of an interface unit, the data acquisition card installed in the computer, and the necessary software for recording the signals generated by the sensors and transfer them to the computer. Two vessels were used (Fig. 2) as sample and reference vessels respectively. The sample, whose aroma is to be detected, was placed in the sample vessel. The hot distilled water, the same as was used for preparing the test samples, was used as the reference vapour in the reference vessel. The reference vessel is used to provide a stable baseline signal and it also reduces the chances of interference from unwanted aroma [13]. Thin plastic tubes were

connected from the vessels to the EN chamber to transfer the aroma profile into contact with the gas sensors. The interface unit controls a valve, which allows the odour molecules to pass in accordance with the user's requirement. A diaphragm pump (vacuum pump manufacturing Co Ltd, UK) was used to facilitate sampling the headspace of the vessels. Throughout the EN tea aroma sniffing process the tea aroma of each tea sample was injected in turn into the headspace of our EN's sensors for 20 complete cycles. The duration of each cycle was set to be 600 s. There were 20 other sniffing cycles interspaced during which the vapour from the reference vessel is injected into the sensor headspace. These cycles were also of 600 s duration. The pumping frequency was set with a sampling time of 1 s for the whole experiment. Sniffing
roceeds in sequence as follows:

(i) Injection of vapour into the sensor headspace from the reference vessel.
(ii) Injection of tea aroma into sensor headspace from the sample vessel.
(iii) Repetition of the processes (i) and (ii) for 20 cycles each.

This means the sniffing process continued sequentially and the pump injected the vapour from either the sample vessel or reference vessel alternately one at a time. We had found experimentally that pumping from reference vessel vapour for 600 s duration is sufficient to allow the sensors to return to their baseline between two consecutive cycles of tea aroma sniffing. This ensures that the EN system only responds to the tea aroma rather than any residual smell from the surroundings. Figure 3 shows a response curve obtained using LabVIEW software (National Instruments Inc.) during EN sniffing of tea samples. It can be observed that the curve increases exponentially during the sniffing of the tea samples and then decreases to a level commensurate with the level of response determined by the odour in the reference vessel. The sensor responses were continuously monitored and stored in data files for the purpose of subsequent data processing off line. The resulting dataset comprised 260 recordings.

Data analysis

Two different approaches were adopted for the data analysis. Firstly, algorithmic or parametric methods based on mathematical analysis requiring a complete knowledge of the system (for example PCA [12]) were adopted. These techniques require the intervention of the experts. Secondly, Intelligent Systems approaches were adopted to process the data based on knowledge of the input parameters; and sometimes the desired corresponding output parameters. In both the cases, the objective was to identify an odorant sample and to estimate the concentration of the aroma. These two steps were further subdivided into four sequential stages: preprocessing, feature extraction, classification, and decision-making; described as follows:

Data preprocessing

There are two main purposes in this stage: (1) to compensate for any sensor drift which typically occurs during data gathering and (2) to account for the transient response of the sensors to the tea aroma, as well as to the reference odour, have to be accounted for in order to optimize the performance of the system. The preprocessing stage also has the effect of compressing the transient response of the sensor array and reducing sample to sample variations. A difference model was used to compensate for example for the effect of drift etc. in the signal. All of the EN tea data was normalized by subtracting the base line of the transient signal. Two different characteristics of the response curves were selected for this purpose. These were the 'peak' and 'mean' of the transient signal.

Feature extraction

Feature extraction had two main purposes: (1) to reduce the dimensionality of the measurement space, and (2) to extract relevant information for 'pattern recognition'. This dimensionality reduction stage projects the initial feature vector onto a lower dimensional space in order to avoid potential problems associated with high-dimensionality, sparse datasets and so on. Moreover, optimum feature extraction removes a major portion of redundant data, which may be perceived as noise in the signal. The resulting low-dimensional feature vector was then used to facilitate classification of the data.

Data clustering

The data were visualized by using data clustering techniques. These allow any clusters or groupings in the data to be identified. We applied three such techniques; namely Principal Component Analysis (PCA), K-means clustering, and Kohonen's Self Organising feature Map (SOM). Let us consider each in turn:

Principal Component Analysis (PCA) is a technique which is very often used for data visualization. It seeks to reduce the vector dimension of the dataset and thus makes it possible to identify the most important, or the principal, components. PCA was shown to be a useful method for visualizing any patterns existing in the response of a multisensory array data so as to facilitate the detection of vapours and odours [12]. In this work, PCA was used to investigate how the response vectors from the sensor array cluster vary in multi-sensor space. PCA analysis should help identify any simple categories for the different tea aroma relating to different tea grades.

The K-means clustering algorithm [15] affords a method of estimating the number of heterogeneous clusters present in a given dataset by iteratively adding and adjusting cluster centers. The algorithm calculates the sum-squared distances of each data point to its cluster center and therefore provides a measure of the error associated with each cluster configuration.

Kohonen's Self Organizing feature Map (SOM) is a competitive and unsupervised learning technique [16]. The basic principle of SOM is to map the input data patterns onto an 'N dimensional grid' of neurons or units. By definition, a SOM is a network formed by the N neurons arranged as the nodes of a planner grid; so that each neuron has four immediate neighbors. The grid forms what is known as the 'output space', as a response to the 'input space'; where the data patterns are presented. This mapping tries to preserve topological relationships. That is patterns that are close in the input space will be mapped to units that are close in the output space, and vice-versa. So as to allow for easy visualization, the output space is usually 1 or 2 dimensional. A SOM network is able to accumulate statistical information about the data without any supplementary information other than that provided, in our case, by the EN sensors.

Data classification

During the classification stage the system was trained to identify the patterns that were representative of each odour. When presented with an unidentified aroma, the classification stage was able to assign to it a class label (identify the odorant) by comparing its pattern with those compiled during training. The decision making stage utilized the knowledge gathered during training.

Three different NN structures namely MLP, RBF, and CPNN were adopted for this stage of data classification. The MLP learns by supervision, during the training phase it is presented with training vectors together with the associated targets corresponding to the specific tea aromas. It learns from the input data by adjusting the weights in the network using its specific learning algorithm; many different training algorithms exist that can be applied to the MLP, but the most commonly used algorithm is error backpropagation [17]. The purpose of this algorithm is to minimise the difference between the generated network output and the desired output; termed the error.

The RBF network has been shown to be an efficient approach for interpolating scattered data and has been applied in various fields [18]. It has a similar architecture to the MLP, exhibiting fully inter-connected layers. It differs structurally from the MLP in that the hidden layer employs a different type of neuron, called the Radial Basis (RB) neuron. Like the MLP, RBF also adopts the supervised learning method, being presented with the input patterns and the associated targets. The approach we adopt here is the feed forward connectionist architectures consisting of 10 hidden neurons of radial kernels and an output layer of linear neurons. Each hidden neuron in a RBF is tuned to respond to a local region of feature space by means of a RB function such as the Gaussian. Then the network performs the weight adaptation required by the output layer. The RBF implementation mainly differs from the MLP in the choice of the heuristic used for selecting the basis function centers and width. Hence the radial basis functions form Gaussian component densities that are used in conjunction with a Gaussian mixture model (GMM). The result of this GMM is a system that attempts to estimate the probability density functions.

The CPNN [19] is essentially a PNN (Probabilistic Neural Network) that is grown by sequential addition of the neurons in the hidden and output layers. The neurons are added in response to patterns presented in the training dataset. Prior to adding neurons, an assessment is firstly made as to whether or not existing neurons can perform the function required by the current input. If they can, then they are adjusted to encompass the new training pattern, otherwise a new neuron is added. Berthold and Diamond [19], found that the CPNN algorithm can offer substantial advantages in terms of network size and generalization capability. So, in comparison, whilst the MLP demands a higher computational overhead, the CPNN is inherently computationally a lightweight NN. In this context, the

CPNN learning algorithm offers an attractive framework for the incremental construction of near-minimal neural network architectures.

Results

We will now present results in the five main sections as follows:

Signal pre-processing and feature extraction

Figure 4 shows a cycle of EN sniffing (a complete cycle of sensor response) of tea aroma along with typical variations of the signal due to sensor drift and transient response. The thick solid line shows the acquired signal and the thin solid line represents the signal variations mainly due to drift. If R is the signal response then the difference model reveals that the actual signal is: DR = R_{DE} - R_{BF}, see Fig. 4. Any difference between the sensor response to the sample vessel odour and the reference vessel odour was thus accounted for. So, the peak signal level was measured as the difference between the signal's peak and its base line (signal maximum, line CD, as shown in Fig. 4); the mean was calculated as the difference between the mean transient response and the signal's base line. As discussed previously the mean and the peak were the only feature vectors which were used in the data analysis. So, the length of the feature vector is eight for the four sensors used in the EN system.

PCA, K-means, and SOM

In the data visualization, the PCA analysis resulted in three principal components for each of the vectors. The results of the PCA for the signals from the 195 different samples of 13 different categories of tea samples; using the data normalized as described above, to form our tea dataset are shown in Fig. 5 in 3 dimensions (3D). It can be observed that some samples are easily separated from the rest and form distinct clusters. But some other samples are not easily separable and overlap with other samples. Such phenomena are likely to be due to the variations of differentiability which are characteristic of the aroma of these tea samples. For example, in the experiment, it was observed that the aroma of samples BP (Geleky), BP (F) (Geleky), PD (Comaibund), BP (Comaibund), BOP (Tamulbari), and BOP (Comaibund) and dust (Comaibund) have distinct discrepancy to each other. On the other hand, samples BOPSM (Tamulbari), PF (Tamulbari), PF (Comaibund), BPS (Comaibund), PD (Tamulbari) and BPS (Tamulbari) were not easily separable. The plots show that the aroma of PF (Comaibund) and BPS (Comaibund) overlap each other. Similarly, BOPSM (Tamulbari) and PF (Tamulbari) also overlap. Such outcomes reveal the performance of the EN in finding both the similarity as well as dissimilarity of aroma profiles for the different tea grades.

The EN dataset clusters are visualized in Fig. 6 in terms of a silhouette plot. The plot shows the K-means clustering of 195 different samples of EN tea data. The plot displays a measure of how close each point in one cluster is to points in the neighboring clusters. It can be observed from the plot that some samples are misclassified in the thirteen classes of aroma. For example, cluster number 5 contains the most numbers of samples, whereas cluster number 10 contains the lowest numbers of samples. Beside the data clustering, the K-means method was particularly efficient at estimating the number of basis function neurons that are required in the hidden layer of a RBF network. Figure 7 shows the error associated with a number of cluster configurations for this dataset. As is evident from the Fig. 7, the optimal number of clusters appears to be between 10 and 20.

For SOM based data clustering, the algorithm was found to be very useful in clustering EN data as it does not rely on any previous knowledge about the class membership of the data. Figure 8 represents the surface of the codebook generated by the SOM during training with the EN dataset of the 195 different samples. It can be seen that the SOM network has clustered the dataset into 8 (eight) distinct clusters instead of 13 (thirteen) different categories.

Neural networks

It was shown in the previous section that our EN data can be successfully clustered by PCA, K-means and, more accurately, by SOM training. That is the SOM result suggested that although the dataset resulted from sniffing 13 different samples, only 8 varieties of aroma appeared to be present. Let us now consider how the NNs performed:

Two MLP networks (the number of layers is set to 2) were tested by considering sample classes as either 13 or 8; one at a time. The first network was required to map the 8 input neurons (features) to 13 output neurons (8::13 network) using Bayesian Regulation back propagation. Whereas, the second network was required to map the same 8 input neurons to 8 output neurons (8::8 network). The second network was based on the evidence from the results of tea tasters, PCA and SOM clustering. The weights were trained with the error back propagation algorithm. The activation functions for the neurons in the hidden layers
(8 neurons, in both cases) employed the sigmoid function. It was observed that the network had very low computational
complexity as training took less than 5 min using a 3 GHz CPU. While working with 195 training samples and 65 testing samples, the 8::13 network resulted in 61.54% classification accuracy. On the other hand, the 8::8 network outperformed it with 90.77% classification accuracy on the same dataset.

The RBF was trained using a hybrid algorithm that employs unsupervised learning for the hidden layer followed by supervised learning for the output layer. The RBF was trained and tested using the datasets used for previous MLP. It was observed that the RBF was able to classify the test sample with 92.31% accuracy.

The training and testing of the CPNN was performed to depict the knowledge of 8 data clusters in the EN tea data. It was observed that the network created 29 hidden neurons (out of 195 maximum, the number of training samples) during training. It was developed, as with the other ANN, with 195 training samples and 65 test samples. It resulted in 93.85% classification accuracy.

Discussions

This paper has reported on work concerned with the analysis of tea aroma in terms of the qualitative and quantitative approaches using a NN based EN system. Tea aroma is one of the three flavour attributes, along with taste and astringency of tea, which are usually considered by tea tasters for tea quality assessment. As one would expect although tea quality is evaluated based on a number of parameters, an attractive aroma is essential for good quality tea. Indeed, in most cases the composition and concentration of the aroma compounds has been shown to play an important role in the valuation and pricing of tea. The main aim of this work was to explore efficient ways via which to investigate the processes for monitoring the quality of tea in terms of its aroma. In this context, an EN with four different metal oxide semiconductor sensors was used to sniff the aroma of thirteen different grades of tea. The transient responses produced by the sensors were stored for subsequent analysis. The mean and peak of the transient responses of the four sensors were extracted as the feature vectors for each of the samples. Our results showed that PCA data visualization indicates that there are eight distinctly separable regions in the dataset. K-means clustering was able to cluster them into thirteen different clusters with some misclassifications. Finally, the SOM clustering suggested that the dataset contained only eight different clusters; this result was comparable to the PCA plot where the principal components were placed in approximately eight different places in a 3D plot. This conclusion was similar to that drawn from the tea tasters' reports, where samples were categorized into seven different classes. The dataset was further classified using several NN based techniques. The MLP resulted in 61.54% accuracy in classifying the dataset into thirteen different categories; but it achieved 90.77% while trying to classify the dataset into eight different categories. The RBF and CPNN achieved 92.31 and 93.85% accuracy respectively. From these results, it can be concluded that tea aroma analysis can be realised using a NN based EN system. Hence such an approach should have a key role to play in other similar areas of application.

Acknowledgements The authors wish to thank the Commonwealth Scholarship Commission in London for their funding, and support for S. Borah, without which the work would not have been possible.

References

1. P.K. Mahanta, P. Tamuli, L.P. Bhuyan, J. Agric. Food Chem. 41, 1677–1683 (1993)
2. H. Horie, S. Fukatsu, T. Mukai, T.G.M. Kawanaka, T. Shimohara, Sens. Actuators B Chem. 13(1–3), 451–454 (1993)
3. N. Togari, A. Kobayashi, T. Aishim, Food Res. Int. 28(5), 495– 502 (1995)
4. J.A. Wright, M. Fishwick, Phytochemistry 18, 1511–1513 (1979)
5. T. Yamanishi, in Tea, Coffee and Cocoa and Other Beverages, ed. by R. Teranishi, R.A. Flath, H. Sigisawa. Flavour Research (Marcell Dekker, New York 1981), pp. 231–304
6. T. Takeo, P.K. Mahanta, J. Sci. Food Agric. 34, 307–310 (1983)
7. P.O. Owuor, Food Chem. 45, 119–124 (1992)

8. R. Ravichandran, R. Parthiban, Food Chem. 62(3), 347–353 (1998)
9. R. Ravichandran, Food Chem. 84(1), 7–11 (2004)
10. M. Bhuyan, S. Borah, in Use of Electronic Nose in Tea Industry. Proc. Int. Conf. EAIT (IIT, Kharagpur, Dec 10–12, 2001), pp. 848–853
11. P.N. Bartlett, J.M. Elliot, J.W. Gardner, Food Technol. 51, 44–48 (1997)
12. J.W. Gardner, Sens. Actuators B Chem. 4, 109–116 (1991)
13. J.W. Gardner, P.N. Bartlett, Sens. Actuators B Chem. 18(1–3), 210–220 (1994)
14. E.L. Hines, E. Llobet, J.W. Gardner, IEE Proc. Circuit Devices Syst. 146(6), 297–310 (1999)
15. J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms (Plenum Press 1981)
16. T. Kohonen, Proc. IEEE (special issue on Neural Network) 78(9), 1464–1480 (1990)
17. D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature 323, 533– 536 (1986)
18. J. Moody, C. Darken, Neural Comput. 1(2), 281–294 (1989)
19. M.J. Berthold, J. Diamond, Neurocomputing 19(1–3), 167–183 (1998)

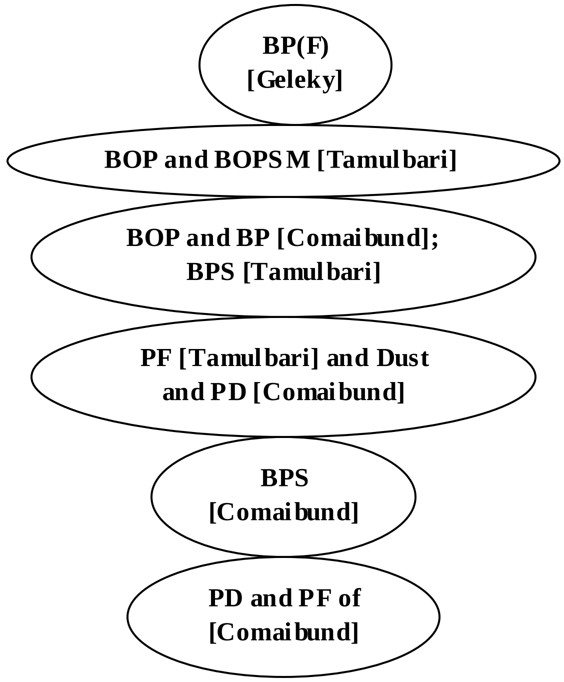

Fig. 1 Tea tasters' category of the selected tea grades for the EN experiment

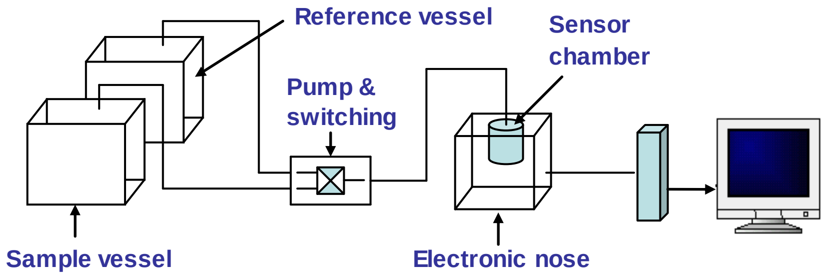

Fig. 2 Experimental setup of the EN

Table 1 The list of SnO2 sensors along with their manufacturer and sensitivity.

Sensors		TGS 880	TGS 826	TGS 825	TGS 822
Main characteristics	**Sensitive to**	Cooking vapours	Toxic gases NH_3	Toxic gases H_2S	Organic solvents. Alcohol, toluene, xylene, etc
	Range (ppm)	10 to 1000	30 to 300	5 to 100	50 to 5000

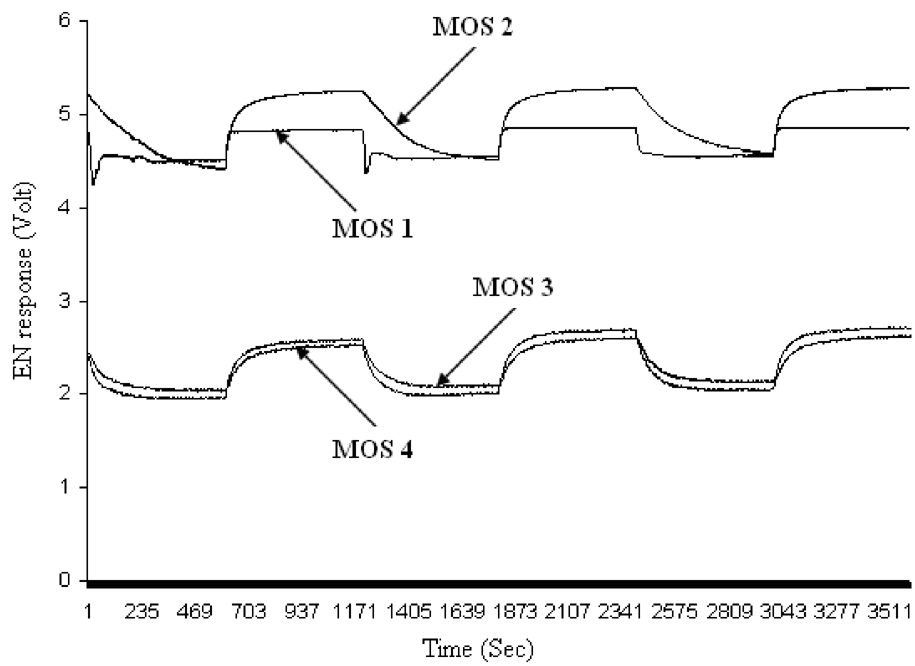

Fig. 3 Typical EN sensors response curve

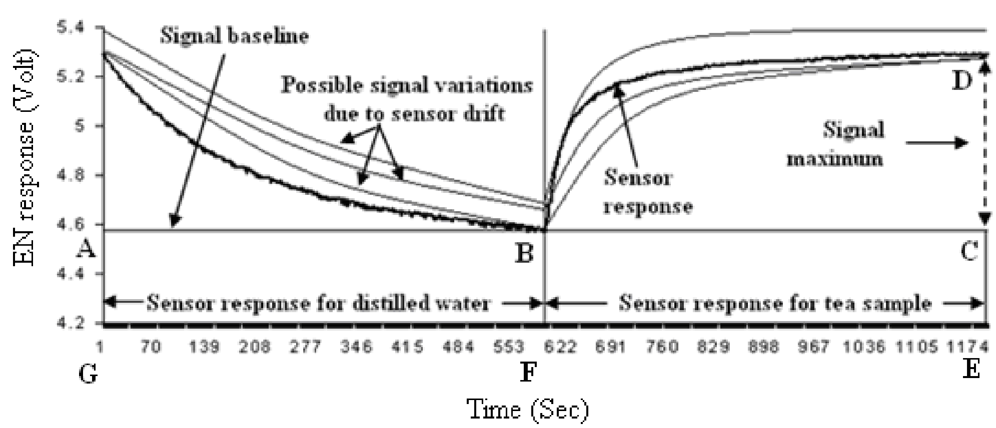

Fig. 4 Sensor response and variations due to sensor drift

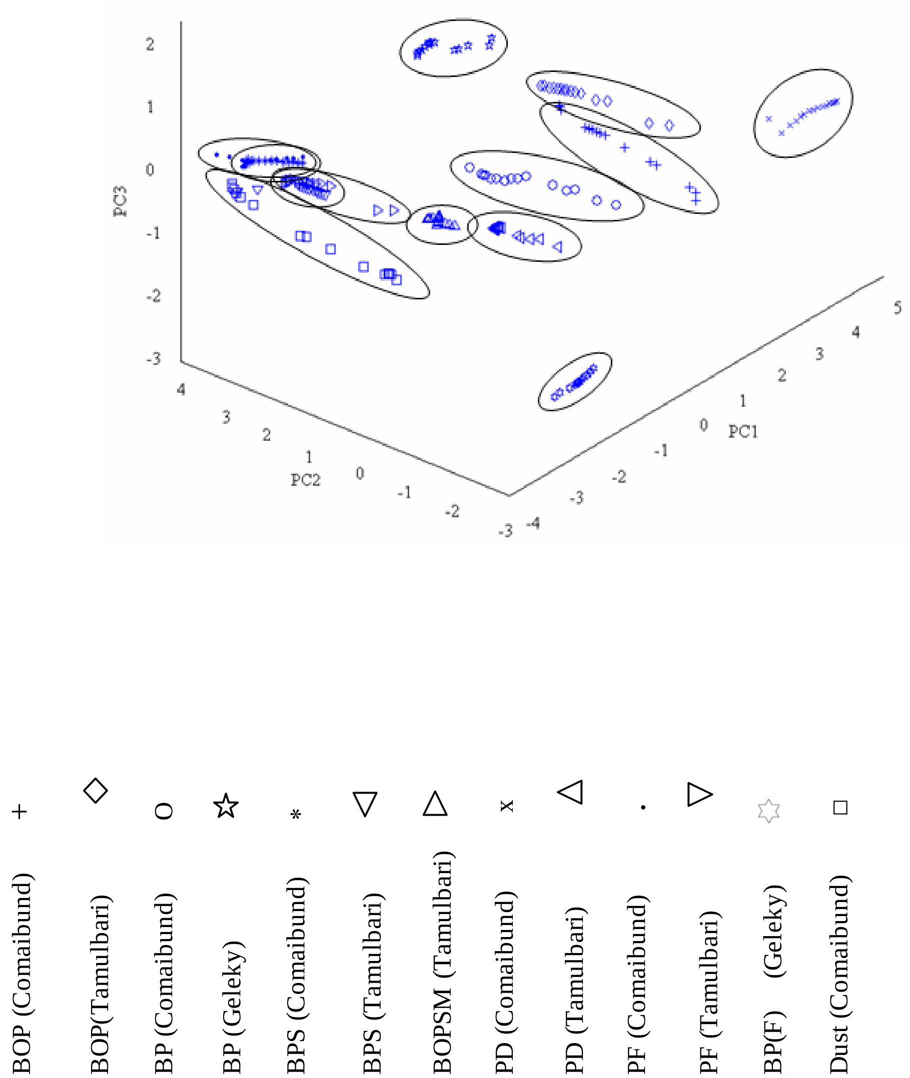

Fig. 5 PCA plots of the EN data for thirteen tea samples and corresponding symbols

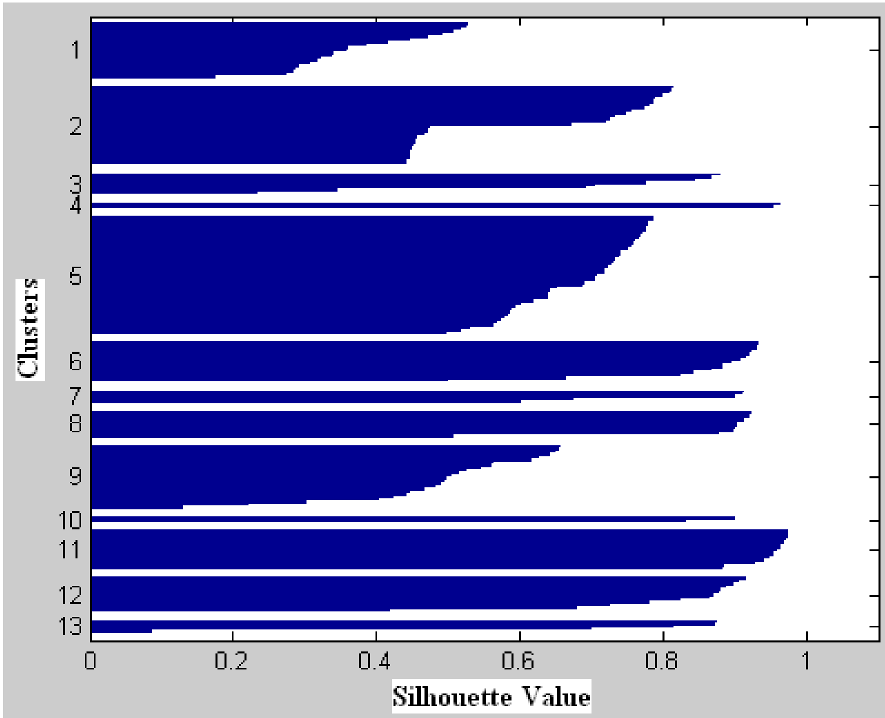

Fig. 6 K-means clustering Silhouette plot

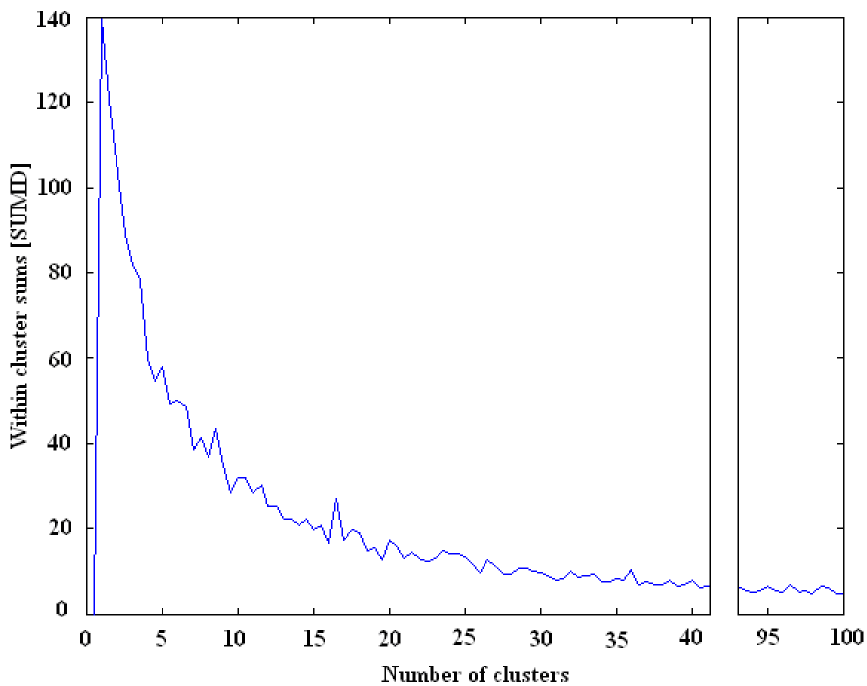

Fig. 7 Error associated with the number of cluster configurations based on K-mean

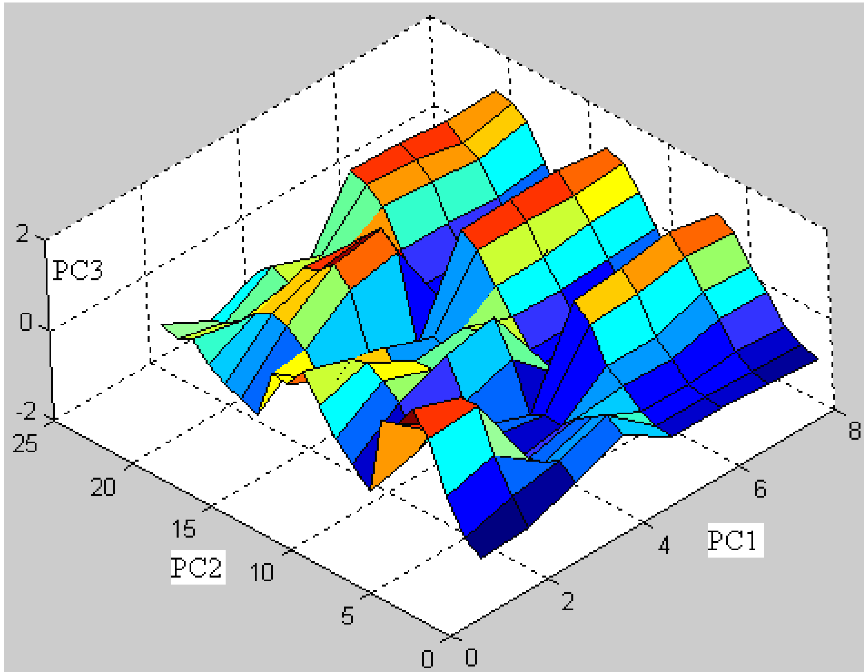

Fig. 8 Surf representation of SOM codebook