Neural network-based electronic nose for cocoa beans quality assessment

Olunloyo, V. O. S.[1], Ibidapo, T. A.[1], R. R. Dinrifo[2]
(1. *Department of Systems Engineering, University of Lagos, Akoka, Nigeria*;
2. *Department of Agric Engineering, Lagos State Polytechnic, Ikorodu, Nigeria*)

Abstract: In this study, a prototype electronic nose was developed for monitoring the quality of cocoa beans. The system comprises an array of metal-oxide semiconductor sensors and an artificial neural network pattern recognition unit. The results obtained from assessment experiments on cocoa beans show good agreement with those obtained from the traditional 'cut test', recording up to 95% accuracy. This investigation demonstrates that the electronic nose technique holds promise as a successful technique in evaluating the quality of cocoa beans for industrial processing.

Keywords: cocoa, quality, electronic nose, neural networks, smell, food

1 Introduction

Earlier approaches to assuring food quality based only on end product testing are no longer considered adequate to ensure food safety in today's circumstances. New food quality control paradigms now focus on ensuring quality at every stage and preventing food hazard throughout the food chain. Thus, much emphasis is currently placed on monitoring the quality of agricultural raw materials that go into food processing.

In many food industries the qualitative evaluation of products is strictly related to human perception of odours (Gomez et al., 2006a). Electronic odour sensing systems seem poised to become a useful tool in complementing or elaborating this important quality control function. The term "electronic nose" or "e-nose" has been used to describe an instrument consisting of an array of weakly specific or broad-spectrum chemical sensors that intend to mimic the human olfactory system, by converting sensor signals to data that are then analysed with appropriate software (Gardner and Bartlett, 1994; Munoz, Steinthal and Sunshine, 1999). Currently, the biggest market for electronic noses is the food industry (Pisanelli et al., 1994). Numerous applications have been reported both in food production and distribution. These include regulating food cooking processes (Almeida, 1994), inspection of fish (Figen et al., 2001; Chantarachoti et al., 2006; Diego and Balaban, 1998), monitoring the fermentation process for wine and in brewing (Weber and Poling, 1997; Pinheiro et al., 2002; Lamagna, 2005); checking fruit juice quality (Young et al., 1999; Gardner et al., 1994; Benedetti et al., 2004; Bleibaum, 2002); classification of coffee aroma, monitoring coffee ripening (Fukunaga et al., 2000; Singh et al., 1996; Falasconi et al., 2005); classification of honey (Benedetti et al., 2004), determining maturity and ripeness of fruits (Hines,1999; Wang, 2004; Maul et al., 1998; Gomez, et al 2006a; Gomez et al., 2006b); grain quality assessment (Borjesson et al.,1996; Miller, 2000); tea quality assessment (Borah et al., 2008); olive oil grading (Christian et al., 2002; Guadarrama et al., 2001; Stella et al., 2000; Guadarrama et al., 2000); detection of adulteration of edible oil (Zheng et al., 2006); quality control of medicinal plants (Baby et al., 2005) and of alcoholic beverages (Pilar et al., 2005), etc.

In recent times, the quality of cocoa beans has become a source of concern in the international cocoa trade. In fact, importers have been requesting for a review of current cocoa quality standards and test methods. There are two classes of quality assessment tests for cocoa. The first is a set of tests carried out on the raw beans while the second series comprises of those tests carried out by the end users. For the dried cocoa beans, the sampler selects at random a significant percentage of the bags for inspection and a stabbing iron is used to pick a number of beans from the selected bags. If the cocoa is in bulk, samples are taken at random from the beans as they enter a hopper or as they are spread on tarpaulins.

Internationally, revisions have been proposed to cocoa beans quality standards and test methods to take cognisance of the views and attitudes of both consumers and processors such that they satisfy both current and forecast needs. The Association for Chocolate, Biscuit & Confectionery Industry of the EU (CAOBISCO) for instance, insists that the current test methods for cocoa beans are inappropriate in meeting the desirable standards necessary to develop the entire cocoa industry. It is necessary for exporting countries as well as cocoa consuming countries to participate in developing cocoa quality assessment procedures, which are robust from both a scientific and commercial viewpoints.

Hashim and Plumas (1999) reported a study on the discrimination of cocoa beans flavour when roasted at different temperatures. Their work was aimed at finding the optimum temperatures for roasting the cocoa beans, rather than on the quality of the raw beans. Alpha MOS (an e-nose system manufacturer) had also reported limited tests on dried raw cocoa beans samples, using the FOX 4000 Sensor Array System (www.alphamos.com). However, it is important to note that cocoa flavour is strongly defined by the origin of the raw cocoa beans (Frauendorf and Schieberle, 2006) which could be traced in turn to how the raw cocoa was processed (Faborode and Dinrifo, 1994; Faborode *et al*, 1995). To avoid a low-quality product and to reduce defects during the production process, it is desirable to detect defects at an early stage of cocoa processing and to initiate remedial action as fast as possible. This research explores the applicability of the electronic nose as an objective tool for assessing the conditions of the cocoa beans – the raw materials used in the food industry. Specifically, this study aimed at developing and applying the electronic nose for the assessment of the quality of cocoa beans of Nigerian origin.

2 Materials and methods

2.1 The electronic nose for the assessment of cocoa beans quality

The electronic nose used in this study was developed at the Systems Engineering Department of the University of Lagos, Nigeria. It was designed to have a fully integrated sampling system with the sensor array and hardware controlled by proprietary Windows- based software, developed in LABVIEW environment. The program which was called 'LAGNOSE' controls the operations of the components of the e-nose system such as the sensors, the solenoid valves and the pump, and it is capable of data collection and processing. On the integrated user's interface, the user can easily specify sampling parameters. Thereafter, the instrument controls the 'sniff' procedure and the output data without further input from the user. The set up makes provision for the user to select which of the installed sensors are to be used in the array, and thereafter collate and analyse data from all or only some of the sensors depending on the samples being observed.

The program has a module that organises the data in a suitable format into a file, after which it undergoes data pre-processing and analysis. From there, the data is exported to another module, which performs the pattern recognition. The pattern recognition unit incorporates a MATLAB© node for configuration and training of the artificial neural network. The electronic nose for our problem is based on the scheme illustrated in Figure 1 and implemented as is shown in Figure 2.

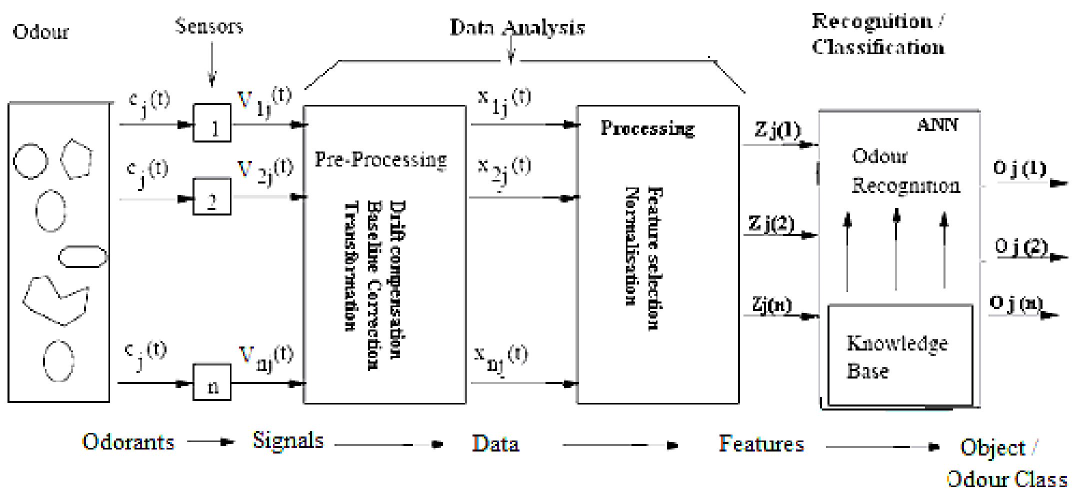


Figure 1 Block diagram of the electronic nose

DAQ: data acquisition card | SC: signals conditioner | P: vacuum pump
Sa: sample chamber | Sc: sensor chamber | SV: solenoid valve
HC: heating chamber.

Figure 2 Electronic nose set up

The sensing element employed in this work is a Figaro TGS (Figaro website: www.figaro.com) gas sensor, which consists of a Tin Oxide (SnO_2) semiconductor, with low conductivity in clean air. The compounds responsible for the cocoa aroma are methylpyrazines, which are nitrogen heterocyclic substances (Perego et al., 2004; Lee, 2006; Bjorn et al., 2005; Bailey et al., 2006) identified the volatile compounds in the aroma of five varieties of roasted and unroasted (raw) cocoa beans as isovaleraldehyde, isobutyraldehyde, propionaldehyde, methyl alcohol, acetaldehyde, methyl acetate, n-butyraldehyde, and diacetyl. TGS sensors have partial selectivity, and in this study they were selected based on sensitivity to these chemicals, as indicated in Table 1.

Table 1 Figaro TGS sensors employed for the electronic nose

Sensors	Chemicals to which optimally sensitive
TGS 825	Hydrogen sulphide
TGS 826	Ammonia, amines
TGS 880	Cooking vapours/ethanol
TGS 813	Combustible gases
TGS 823	Alcohols / Organic solvents
TGS 830	Refrigerants, R22/R21

A user friendly user interface was designed. The parameters for the acquisition could be set, and the choice of what kind of operation to be performed could be made at the interface. Figure 3 shows a screenshot of the users' interface.

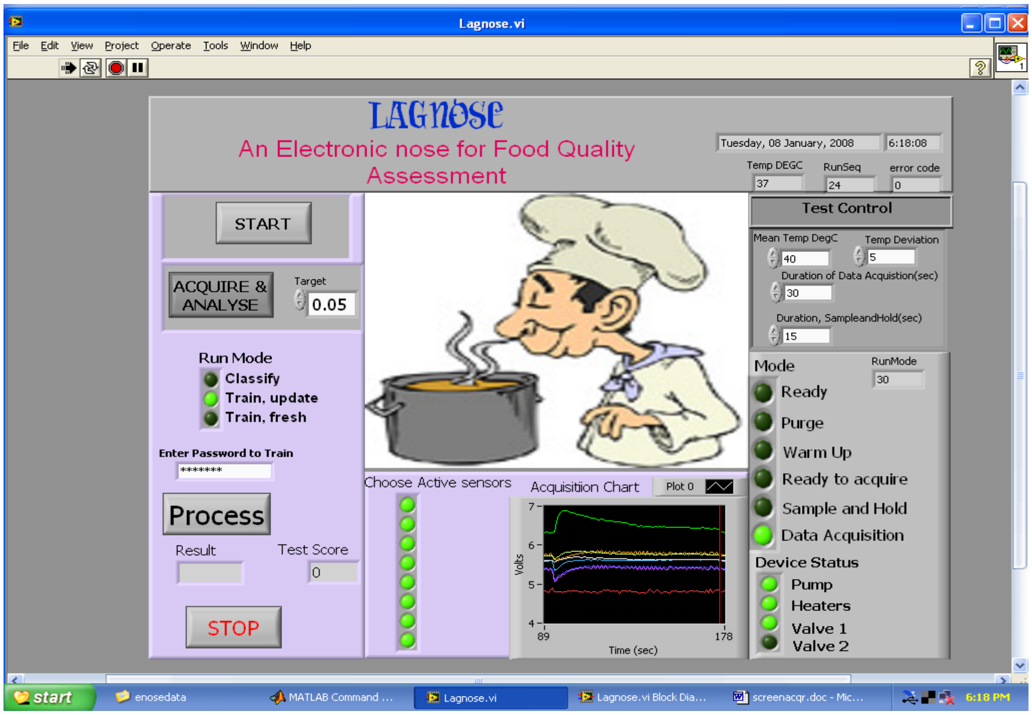


Figure 3 Screen shot of the user interface

2.2 Assessment of cocoa beans quality

The samples of cocoa used for this work were sourced from two cocoa beans handling enterprises namely Cocoa Industries Ltd (CIL), Ikeja, Lagos and Commodities Support Services Ltd (CSSL) Apapa Road, Lagos. CIL has been in the business of cocoa processing for more than 40 years, producing cocoa cake, butter, and cocoa beverages for the local markets and for export. CSSL principally serves as quality certification agents for cocoa exporters in Nigeria. The company serves cocoa beans exporters from all the 14 cocoa producing States in Nigeria. Thus, we were able to collect samples from all the cocoa regions of Nigeria. The cocoa beans samples were classified into different quality classes with the assistance of officers in the Laboratories of these establishments using standard test methods. The moisture content of the raw dried cocoa beans falls in the range 7%-8% approximately.

Working on maize grains, Borjesson et al., (1996) described odour classes for the food grains as normal, musty, mouldy, acid, sour, burnt or foreign, and the intensities of off-odours were given as weak, pronounced, or strong. Also, in another work by Alpha MOS, (www.alphamos.com) only two classes (good, bad) were used for cereals. Thus, it appears that the level of classification used depends on the product and the purpose for which the classification is being made. The principal odour description for cocoa beans includes normal (well fermented - chocolaty smell), mouldy and smoky. Since the present practice in cocoa processing industries is to accept cocoa bean lots based on whether the beans are good (well fermented, good chocolate flavour), or bad (mouldy, or showing insect infestation), we therefore consider it sufficient to classify the beans into these two classes.

2.2.1 Cut tests of the samples used

The cut test provides an assessment of the beans from which analysts may infer certain characteristics of the cocoa, which gives an indication of quality. It is to be noted that the cut test merely determines acceptability of a produce lot based on counting and estimating the percentage of the defects present in it. For the cut test (ISO 1114:1977) states:

> "For the determination, 300 beans shall be opened or cut lengthwise through the middle, so as to expose the maximum cut surface of cotyledons. Both halves of each bean shall be visually examined in full daylight or equivalent artificial light. Each defective type of bean shall be counted separately, and the result for each kind of defect shall be expressed as a percentage of the 300 beans examined".

The defects normally recorded are beans with purple colour (poorly fermented) mould growth, slate beans, insect-damaged, germinated and flat beans

Batches of cocoa bean samples were randomly selected from the warehouse at CSSL facility at Apapa Road Lagos. Using the rules of the cut test, one batch confirmed to be acceptable and another assessed as unacceptable were selected. The test was done with the assistance of laboratory officers at CIL and CSSL. Any lot with more than 5% visibly mouldy beans or 10% slaty beans or more than 10% insect damaged or germinated was regarded as bad or unacceptable. If the percentages of defects were less, the sample is considered to be good.

2.2.2 Cocoa quality by the electronic nose

The sensitivity characteristics of the TGS sensors are altered by changes in atmospheric temperature and humidity. As a result, we used only well dried cocoa beans of moisture content 7%-8% , which is the recommended moisture content for commercial trading of the beans. In addition, the samples to be tested were pre-heated in a heating chamber (see Figure 2) enclosed in a water bath, the temperature of which was controlled by a sensor integrated with the acquisition programme. A humidity sensor was installed inside the sensor chamber. Our experiment for the acquisition of cocoa smell fingerprints took place at sample chamber temperature of 60℃ and relative humidity of 70%.

During the first phase (baseline phase), a carrier gas (filtered air in this case) is blown through the sensor chamber and sensor baseline is acquired. In the second phase (absorption phase), the 3-way valve connects the sensor chamber to the sample chamber. The headspace generated over the sample is then sucked into the chamber, exposing the sensors to the odorant. In the third phase (desorption phase), the valve is switched again to the baseline position and the odour is flushed away by the gas carrier.

3 Results and discussion

3.1 Cut test results

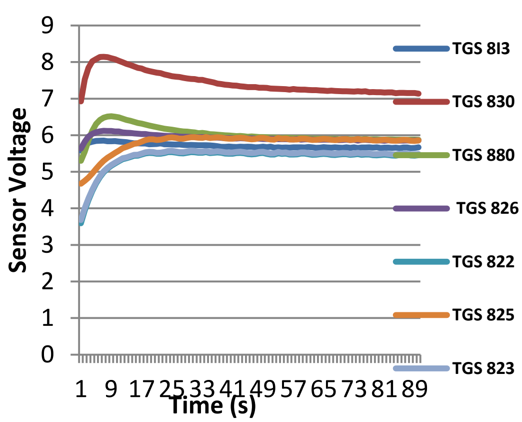
The cut tests were conducted to confirm the condition of the beans to be used for electronic nose experiment. Table 2 shows

Table 2 Cut test results for the cocoa beans batches

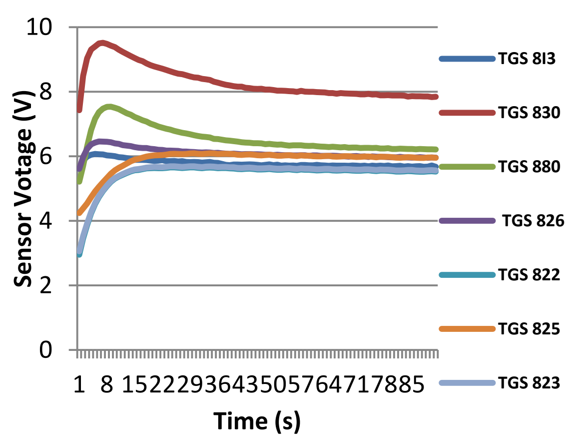
Cocoa beans lot	Mouldy/%	Insect damaged /%	Slaty %	Quality class
A	6	4	3	Unacceptable (bad)
B	1	0	1	Acceptable (good)

3.2 Electronic nose response to cocoa beans aroma

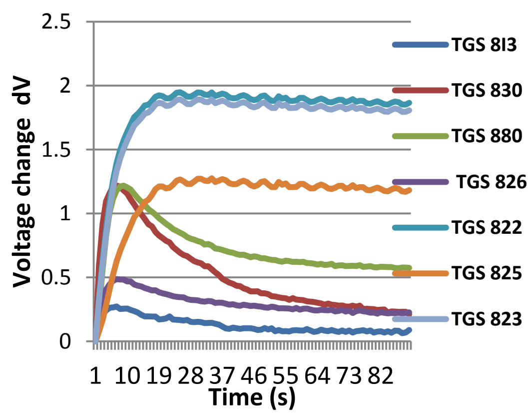
The response of the sensors during the test on a sample of cocoa beans is shown in Figure 4. It shows that, after an initial period of low and stable conductivity (during which only clean air is flowing through the measurement chamber), conductivity increases sharply and reaches a peak after about 25 s.



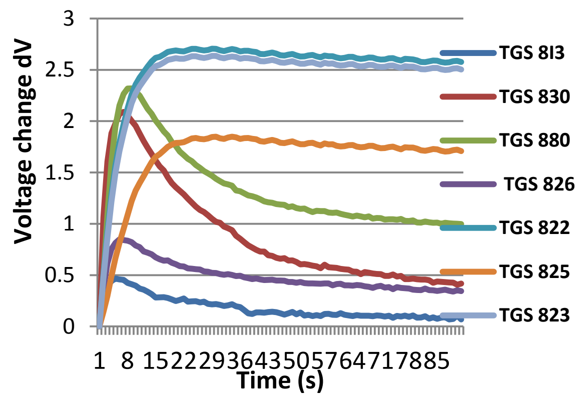
a. Raw sensor output, V–bad beans



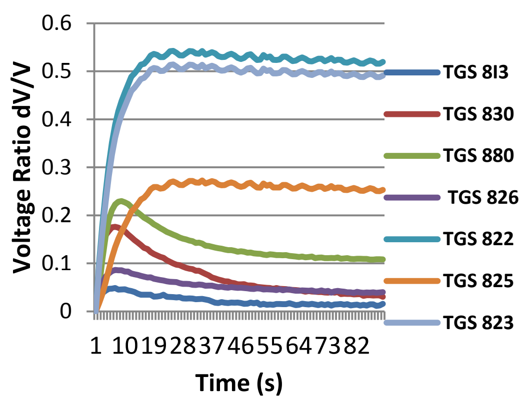
b. Raw sensor output, V-good beans



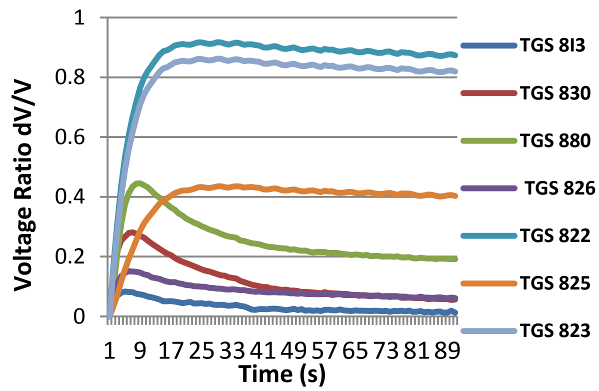
c. Voltage change from baseline,V-badbad beans



d. Voltage change from baseline, V–good beans



e. Voltage ratio (compared to baseline)-bad beans



f. Voltage ratio (compared to baseline) – good beans

Figure 4 Sensor response curves (a and b) and pre-processed response curves (c, d, e and f)

3.2.1 Data exploration

The feature parameters were extracted in accordance with the methods of Roussel et al., (1999) and Zou et al. (2003). The most promising features were identified based on their discrimination power using Distinction Index (D.I). According to Zou and Zhao (2004) and Zou et al. (2003), the D.I can be evaluated as:

$$D.I = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2}$$

where μ_1 and μ_2 are the mean values of the feature parameters calculated from the signals measured in state 1 and state 2 and with σ_1 and σ_2 as their respective standard deviations.

The larger the value of D.I, the better the feature parameter will be as a discrimination criterion.

Several feature parameters were considered, but twelve were ultimately chosen based on tests performed using their D.I scores (Table 3).

Table 3 Distinction indices (D.I) and discriminant factor communalities (DFC) for the best 12 feature parameters selected.

Feature parameters	Peak1	Peak2	Peak3	Peak4	Rten1	Rten2	Rten3	Rten4	Tp1	Tp2	Tp3	Tp4
D.I	1.351	1.104	1.110	0.727	0.714	0.421	0.46	0.634	0.201	0.182	0.158	0.162
DFC	0.409	0.124	1.016	0.510	0.495	0.672	0.326	0.121	0.134	0.135	0.348	0.284

For this work, the back-propagation algorithm was used to train the network, which was then implemented using Neural Networks Toolbox programming utility in MATLAB.

In order to analyse the electronic nose data appropriately, it was necessary to investigate which pre-processing method was the most valid. The methods of pre-processing carried out on all datasets included the following signals baseline correction methods: fractional, relative and difference. Figure 4c, d, e and f show the curves obtained on the baseline corrected data.

3.2.2 Principal component analysis (PCA)

Principal Component Analysis (PCA) was performed on the baseline corrected and normalized data (See table A1 in the appendix). This was done with a view to identifying clusters in the data. Figures 5-8 show the results of the analysis. Table A2 in the appendix shows the influence of baseline correction (data pre-processing) on data clustering.

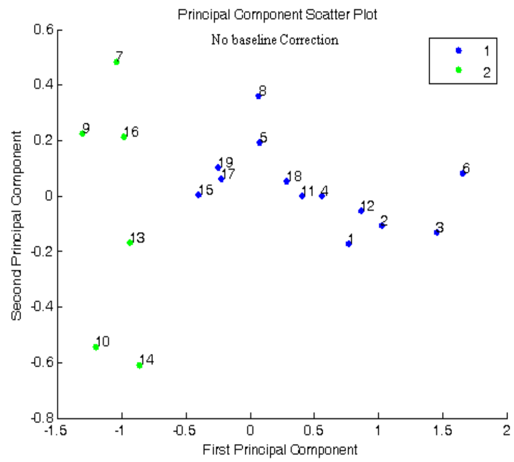


Figure 5 PCA scores
(No baseline correction, normalized data)

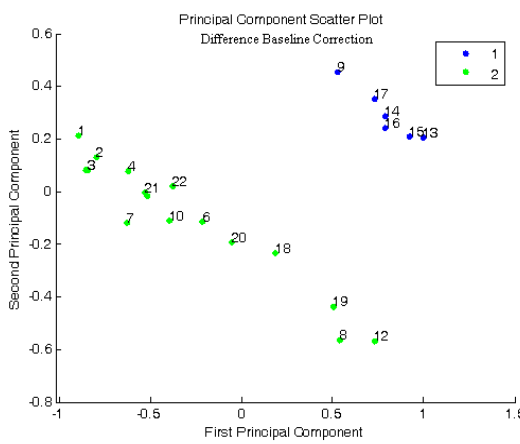


Figure 6 PCA scores
(Difference baseline correction, normalized data)

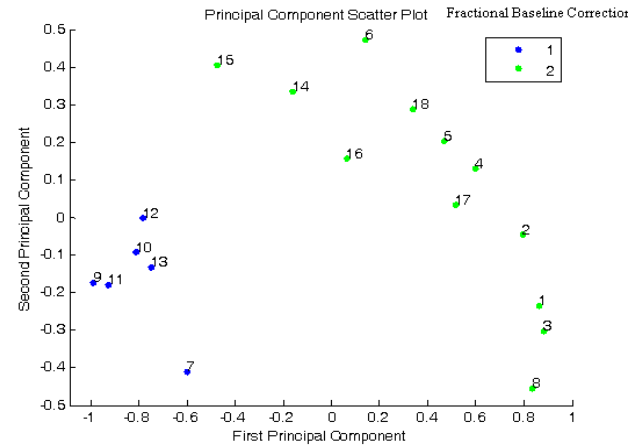


Figure 7 PCA scores
(Fractional baseline correction, normalized data)

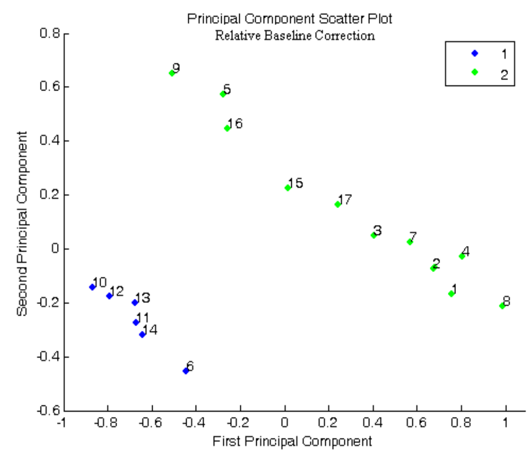


Figure 8 PCA scores
(Relative baseline correction, normalized data)

The analysis revealed that in the case of no baseline correction, it is difficult to obtain clusters as the data in Figure 7 showing a scatter plot, indicating little correlation with classes of the sample. This suggests that baseline correction is important if classification of the data is to be successfully performed. The analysis also shows that while "relative", "difference", and "fractional" baseline correction show some measures of success in clustering and hence discriminating the data, "fractional" method shows a superior result, followed by "relative" and "difference"

3.3 Classification of cocoa beans data using ANN

3.3.1 Choosing the parameters for the ANN

It was necessary to investigate the discriminatory power of each of the features extracted in the earlier section. For this present work we used the Distinction Index (DI) method to determine which of the feature parameter are the most important (feature optimisation). Only four of the sensors show appreciable response as is shown in Figure 3. The twelve (12) features extracted were $Peak_1$, $Peak_2$, $Peak_3$, and $Peak_4$ together with $Rten_1$, $Rten_2$, $Rten_3$ and $Rten_4$, Tp1, Tp_2, Tp_3 and Tp_4 based on their DI values where $Peak_n$ (are the maximum (peak) reading of sensor *n*), $Rten_n$ (the tenth value reading for sensor *n*) and Tp_n (time to reach peak reading).

3.3.2 The artificial neural network (ANN) architecture

To reduce noise, signals from two of the sensors were neglected, as they were always the same, irrespective of the type of cocoa beans that was being sampled. It is to be noted that the fourth parameter, the adsorption slope was also dropped based on the fact that it does not have significant contribution to sample discrimination. The architecture adopted after all due consideration was:

a. Input layer node = 12
b. Number of hidden layers = 1
c. Hidden layer nodes = 24
d. Output layer node = 2

The output layer node indicates the output of the network for every instance of input data presented. The result is either "GOOD" (i.e. [1.00, 0.00] but coded as [0.95, 0.05]) or "BAD" (i.e. [0.00, 1.00] but coded as [0.05, 0.95]).

e. Termination condition: error means square < 10^{-10}
f. The activation function of each node is given by the sigmoid function as follows:

$$\text{Activation function} = \frac{1}{1+e^{-netj}}$$

where *netj* is the activity of node j.

g. Training function: Gradient descent with momentum and adaptive learning rate.

3.3.3 Training and test sets

In order to develop the pattern recognition system, the sample data (collected from 120 experiments) were split into two sets, namely, the training set and the test set. The training set (80 samples) was used to establish the design parameters of the pattern recognition system, while the test (40 samples) set served to evaluate the system performance. Typically, the performance of the pattern recognition system is measured by computing the percentage of correctly recognised patterns on all the input patterns presented to the system. In order to validate the efficacy of the ANN model, we used an additional validation set of 20 samples to test the network.

3.3.4 ANN results

Our ANN results indicate the success with which the program predicted the class of a sample drawn from the lot already classified using the cut test. It is necessary to emphasize that the cut test merely identified the quality class based on the presence of defective beans above the prescribed threshold (like mouldiness of the beans, poorly fermented, insects' presence, etc). The cut test is not based on the usual hedonic scale and hence the result cannot be subjected to the conventional statistical tests. Of the 40 test samples used, 38 samples were correctly classified (representing 95% correct classification rate) into the two classes: good (fruity- flavoured) and bad (off-flavoured) as in Table 4. The assessment was compared with cocoa beans classified using the cut test.

Table 4 ANN classification of cocoa beans

Actual group	Predicted group		Correctly classified/ %
	Bad	Good	
Bad	18	2	90
Good	0	20	100
Overall correct classification			95.8

4 Conclusions

An electronic nose (E-nose) has been designed and implemented for the classification of classify cocoa beans. The performance was enhanced by pre-heating the raw beans to 60 ℃ in a sample chamber. PCA was employed mainly for data visualization, which showed the discrimination of the dataset into distinctly separable points, with the data pre-processed with 'fractional' baseline correction method giving the best discrimination on the data. The datasets were further classified using ANN- based techniques. The results show that 95% of the cocoa beans data can be classified into two quality classes successfully. The research basically establishes the efficacy of the smell sensor techniques for quality monitoring of cocoa beans.

Acknowledgements

The authors wish to thank the University of Lagos Research Council for the funding for this research, without which the work would not have been possible. The support of Cocoa Industries (CIL) Ikeja and Commodity Support Services Ltd Apapa Road Lagos for the supply of cocoa beans samples is appreciated.

References

Almeida, A. 1994. Neural network in oven prevents overcooking. The Institute: 3.

Alpha, MOS. Applications Notes. http://www.alpha-mos.com/notes/an-en.php

Baby, R, M. Cabezas, E. Castro, R. Filip, and N. E. Walsöe de Reca. 2005. Quality control of medicinal plants with an electronic nose. *Sensors and Actuators B: Chemical*, 106(1): 24-28.

Bailey, S. D, D. G. Mitchell, M. L. Bazinet, and C. Weurman. 2006. Studies on the volatile components of different varieties of cocoa beans. *Journal of Food Science,* 27(2): 165–170.

Benedetti, S, S. Mannino, A. G. Sabatini, and G. L. Marcazzan. 2004. Electronic nose and neural network use for the classification of honey. *Apidologie*, 35: 1–6.

Bjorn, T., W. Olthius, and A. van den Berg. 2005. Ammonia sensors and their applications- a review. Sensor and actuators, B 107: 666-677.

Bleibaum, R. N., H. Stone, T. Tan, S. Labreche, E. Saint-Martin and S. Isz. 2002. Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices. *Food Quality and Preference*, 13(6): 409-422.

Borah, S., E. L. Hines, M. S. Leeson, D. D. Iliescu, M. Bhuyan, and J. W. Gardner. 2008. Neural network based electronic nose for classification of tea aroma. *Sens. & Instrumen. Food Qual*, 2: 7–14 DOI 10.1007/s11694-007-9028-7

Borjesson, T., T. Eklov, A. Jonsson, H. Sundgren, and J. Schnurer. 1996. Electronic nose for odour classification of grains. Analytical Techniques and Instrumentation. *American Association of Cereal Chemists*, 73(4): 457-461.

CAOBISCO. 2002. Cocoa Quality. Association of the Chocolate, Biscuit & Confectionery Industries of the EU, Brussels. http://www.caobisco.com

Chantarachoti, J., A. C. M. Oliviera, B. H. Himelboom, C. A. Crapo, D. G. MacLacLan. 2006. Portable electronic nose for detection of spoiling Alaska pink salmon (Oncorhynchus gorbuscha). *Journal of Food Science*, 71 (5):S414–S421.

Cimander, C., M. Carlsson, and C. F. Mandenius. 2002. Sensor fusion for on-line monitoring of yoghurt fermentation. *Journal of Biotechnology,* 99: 237-248.

Diego, A. L, and M. O. Balaban. 1998. Electronic nose odour evaluation of salmon fillets stored at different temperatures. In *Proc. 5th Symposium on Olfaction and Electronic Nose*, Maryland.

Evor, Hines. 1999. Electronic nose sniffs out fresh fruit. http://www.charlie.iit.edu/enos.html. Accessed May 2006.

Faborode, M. O., and P. R. Dinrifo. 1994. A mathematical model of cocoa pod deformation based on Hertz theory. *International Agrophysics*, 1994(8): 403-409.

Faborode, M. O, J. F. Favierh, and A. Omolayo. 1995. On the effects of forced air drying on cocoa quality. *Journal of Food Engineering*, 25(4): 455-472.

Falasconi, M., M. Pardo, G. Sberveglieri, I. Riccò, and A. Bresciani. 2005. The novel EOS835 electronic nose and data analysis for evaluating coffee ripening. *Sensors and Actuators B: Chemical*, 110(1): 73-80.

Figen, K, D. A. Luzuriaga, and M. Balaban. 2001. Quality evaluation of raw and cooked catfish (Ictalurus punctatus) using electronic nose and machine vision. *Journal of Aquatic Food Product Technology*, 10(1).

Frauendorfer, F., and P. Schieberle. 2006. Identification of the key aroma compounds in cocoa powder based on molecular sensory correlation. *Journal of Agricultural and Food Chemistry*, 54(15): 5521-5529.

Fukunaga, T, S. Mori, Y. Nakabayashi, M. Kanda, and K. Ehara. 2002. Application of flavour sensor to coffee. UCC Ueshima Coffee Co., Ltd., Japan Tokyo Institute of Technology, Japan.

Gardner, J. W., and P. N. Bartlett. 1994. A brief history of electronic noses. *Sensors and Actuators B: Chemical*, 18/ 19, 211–220.

Gomez, A. H., G. Hu, J. Wang, and A. G. Pereira. 2006a. Evaluation of tomato maturity by electronic nose. *Computers and electronics in agriculture*, 54(1): 44–52.

Gomez, A. H., G. Hu, J. Wang, and A. G. Pereira. 2006b. Electronic nose technique potential monitoring mandarin maturity. *Sensors and Actuators B*, 113: 347–353.

Guadarrama, A. M., L. Rodriguez-Méndez, J. A. de Saja, J. L. Rios, J. M. Olias. 2000. Array of sensors based on conducting polymers for the quality control of the aroma of the virgin olive oil. *Sensor and Actuators B-Chem*, 69: 276–282.

Guadarrama, A., M. L. Rodriguez-Méndez, C. Sanz, J. L. Rios, and J. A. de Saja. 2001. Electronic nose based on conducting polymers for the quality control of the olive oil aroma discrimination of quality, variety of olive and geographic origin. *Analytica Chimica Acta*, 432(2):283–292.

Hashim, L., and B. Plumas. 1999. Electronic nose for monitoring cocoa beans aroma, In Hurst (ed) Electronic noses and sensor array based systems - design and applications, Technomic Publishing Co., Inc. Lancaster: 296-307.

ISO 1114 (1970) Cocoa Beans -- Cut test. International organisation for standardisation lamagna, A, D. F. Rodríguez, N. N. Scoccola, and S.L. Reich. 2005. Feature extraction methods for the analysis of an electronic nose response. J. Argent. Chem. Soc.93:1-3.

Lee, D. P. 2006. Odour from pig production: its relation to diet. PhD Thesis Wageningen University and Research Centre. Wageningen, the Netherlands

Maul, F., S. A. Sargent, M. O. Balaban, E. A. Baldwin, D. J. Huber, and C. A. Sims. 1998. Aroma volatile profiles from ripe tomatoes are influenced by physiological maturity at harvest: an application of electronic nose technology. TEKTRAN. United States department of agriculture. Agricultural research service.

Munoz, B. C., G. Steinthal, and S. Sunshine. 1999. Conductive polymer-carbon black composites-based sensor arrays for use in an electronic nose. *Sensor Review*, 19(4): 300-305.

Perego, P., B. Fabiano, M. Cavicchioli, and M. Del Borghi. 2004. Cocoa quality and processing. A Study by solid-phase microextraction and gas chromatography analysis of methylpyrazines. *Trans IChemE, Part C, Food and Bioproducts Processing*, 84(C4): 291–297.

Pilar, M., O. Busto, J. Guasch, and R. Boqué. 2005. Electronic noses in the quality control of alcoholic beverages. *Trends in Analytical Chemistry*, 24(1): 57-66.

Pinheiro, C., C. M. Rodrigues, T. Schaofer, and J. G. Crespo. 2002. Monitoring the aroma production during wine-must fermentation with an electronic nose. *Biotechnol Bioeng*, 77(6): 632-40.

Pisanelli, M. A, A. A. Qutob, P. Travers, S. Szysko, and K. C. Persaud. 1994. Applications of olfactory system using a model nose. Nature 299: 352-355. 1994

Roussel, S; G. Forsberg, P. Grenier, and V. Bellon-Maurel. 1999. Optimisation of electronic nose measurements. Part II: Influence of experimental parameters. *Journal of Food Engineering*, 39: 9-15 (1999).

Singh, S., E. L. Hines, J. W. Gardner. 1996. Fuzzy neural computing of coffee and tainted water data from an electronic nose. *Sensors and Actuators*, 30(3): 190-195.

Stella, R., J. N. Barisci, G. Serra, G. G. Wallace, and D. De Rossi. 2000. Characterisation of olive oil by an electronic nose based on conducting polymer sensors. *Sensors and Actuators B-Chemical.*, 63(1-2): 1–9.

Wang J., A. H. Gómez, and A. G. Pereira. 2006. Acoustic impulse response for measuring the firmness of mandarin during storage. *Journal of Food Quality*, 29(4): 392-404.

Wang N., Z. Li, C. Vigneault. 2006. Electronic nose and tongue in food production and processing. *Stewart Post Harvest Review*, 2(4):1-5(5).

Weber. K., and J. Poling. 1997. Applications of an electronic nose in the brewing industry. *Technical quarterly* - Master Brewers Association of the Americas, 34(2): 127-132.

Young, H., K. Rossiter, M. Wang, and M. Miller. 1999. Characterization of royal gala apple aroma using electronic nose technology-potential maturity indicator. *Journal of Agricultural and Food Chemistry*, 47(12): 5173-5177.

Zheng, H., and J. Wang. 2006. Electronic nose and data analysis for detection of maize oil adulteration in sesame oil. *Sensors and Actuators B*, 119: 449–455.

Zou, X.., J. Zhao, S. Wu, and X. Huang. 2003. Vinegar classification based on feature extraction and selection from tin oxide gas sensor array data. *Sensors*, 3(4) 101-109.

Zou, X., and J. Zhao. 2004. Selection of the most efficient sensors for methanol detection in fake alcohol. *Sensors & Transducers Magazine (S&T e-Digest)*, 46(8): 285-295.

Appendix 1 Extracted features (fractional baseline corrected and normalized)

Max1	Max2	Max3	Max4	Max5	Max6	Max7	Tp1	Tp2	Tp3	Tp4	Tp5	Tp6	Tp7	Min1	Min2	Min3	Min4	Min5	Min6
0.883	0.678	0.000	0.640	0.195	0.458	0.604	0.000	0.792	0.271	0.271	0.104	0.110	0.935	0.820	0.000	0.730	0.114	0.110	0.106
0.883	0.678	0.000	0.640	0.195	0.481	0.604	0.000	0.792	0.271	0.271	0.104	0.110	0.935	0.820	0.000	0.730	0.114	0.110	0.106
0.294	0.253	0.140	0.613	0.182	0.445	0.187	0.013	0.827	0.751	0.480	0.361	0.366	0.311	0.380	0.159	0.654	0.340	0.345	0.349
0.287	0.345	0.150	0.603	0.173	0.438	0.511	0.992	0.850	0.000	0.000	0.326	0.400	0.304	0.220	0.180	0.631	0.254	0.246	0.239
0.877	0.494	0.230	0.645	0.171	0.463	0.449	0.941	0.746	0.554	0.175	0.000	0.000	0.912	0.380	0.267	0.713	0.199	0.188	0.178
0.868	0.402	0.240	0.629	0.164	0.458	0.573	0.097	0.746	0.186	0.073	0.000	0.048	0.904	0.600	0.282	0.697	0.023	0.016	0.010
0.310	0.420	0.480	0.673	0.282	0.499	0.000	0.067	0.665	0.209	0.198	0.028	0.028	0.368	0.000	0.598	0.765	0.000	0.000	0.000
0.348	0.477	0.470	0.666	0.288	0.503	0.044	0.067	0.665	0.209	0.209	0.028	0.035	0.398	0.140	0.557	0.758	0.007	0.007	0.007
0.012	0.172	0.320	0.660	0.328	0.566	0.529	0.597	0.815	0.424	0.407	0.271	0.276	0.009	0.400	0.383	0.761	0.322	0.317	0.313
0.000	0.178	0.300	0.667	0.328	0.564	0.529	0.929	0.705	0.458	0.424	0.250	0.248	0.000	0.860	0.347	0.765	0.348	0.342	0.335
0.179	0.603	0.280	0.702	0.224	0.494	0.471	0.508	0.723	0.339	0.333	0.181	0.186	0.199	0.560	0.275	0.784	0.221	0.217	0.212
0.413	0.425	0.720	0.843	0.452	0.678	0.613	0.987	0.121	0.814	0.514	0.417	0.407	0.403	0.480	0.898	0.878	0.443	0.439	0.435
0.431	0.000	0.760	0.833	0.492	0.709	0.760	0.740	0.000	0.831	0.701	0.632	0.635	0.416	0.920	0.879	0.867	0.551	0.551	0.551
0.104	0.092	0.680	0.821	0.487	0.709	0.702	0.929	0.659	0.802	0.627	0.549	0.545	0.063	0.840	0.770	0.843	0.598	0.594	0.591
0.416	0.103	0.600	0.783	0.493	0.714	0.773	0.366	0.410	0.723	0.718	0.681	0.655	0.411	0.500	0.709	0.824	0.558	0.560	0.562
0.936	0.856	0.450	0.000	0.000	0.000	0.720	0.723	0.624	0.633	0.429	0.333	0.331	0.954	0.900	0.513	0.000	0.395	0.389	0.384
0.947	0.822	0.670	0.131	0.020	0.022	0.893	0.958	0.567	0.638	0.497	0.278	0.400	0.950	0.020	0.753	0.150	0.451	0.443	0.435
0.932	0.851	0.520	0.091	0.052	0.100	0.773	0.458	0.462	0.475	0.311	0.222	0.159	0.943	0.760	0.698	0.171	0.960	0.963	0.963
0.931	0.713	0.230	0.787	0.255	0.273	0.747	0.647	0.786	0.842	0.633	0.549	0.552	0.943	0.840	0.232	0.802	0.968	0.968	0.970
0.017	0.603	0.280	0.702	0.224	0.494	0.471	0.508	0.723	0.339	0.333	0.181	0.186	0.024	0.560	0.275	0.784	0.966	0.978	0.969

Appendix 2 Influence of data pre-processing technique on data clustering by PCA

Actual good class	Actual bad class	Classified as good by " Difference " method	Classified as good by "Difference " method	Classified as good by "Fractional " method	Classified as bad "Fractional " method	No "Pre-processing" classified good	No "Pre-processing " classified as bad	Classified as good "Relative" method	Classified as good "Relative" method
1 2	9	1 2	9	1 2 3	* 7 9	1 2	10	1 2	*6 10
3 4	10	3 4	13	4 5 6	10 11	3 4	*14	3 4	11
5 6	11	5 6	*14	8 14 15	12 13	5 6 7		5 7	12
7 8	12	7 8	*15	16 17		8 *9 *11 12		8 *9	13
14 15	13	*10 *11	*16	18		13 *15 *16		15	*14
16 17		*12 18	*17			*17 *18		16 17	
18								18	

Notes: *****Samples wrongly classified.

"Difference" baseline correction: $S_r = R_i - R_0$

"Fractional" baseline correction: $S_r = \frac{R_i - R_0}{R_0}$

"Relative" baseline: $S_r = \frac{R_i}{R_0}$

Where S_r= corrected sensor response

R_i= response of sensor at time

R_0= response of sensor at the start of signals acquisition

$$\text{"Fractional" baseline correction}: \ \mathrm{S_r} = \frac{\mathrm{R_i} - \mathrm{R_o}}{\mathrm{R_o}} \text{"Relative" baseline}: \ \mathrm{S_r} = \frac{\mathrm{R_i}}{\mathrm{R_o}} \text{where } \mathrm{S_r} = \text{corrected sensor response} \mathrm{R_i} = \text{response of sensor at time i} \mathrm{R_o} = \text{response of sensor at the start of signals aquistion}$$