Temperature Compensated Electronic Nose for Fruit Ripeness Determination Using Component Correction Principal Component Analysis

John Victor M. Lim[1*], Noel B. Linsangan[1], Febus Reidj G. Cruz[1,2], Wen-Yaw Chung[2]
[1] Mapua Institute of Technology, Manila, Philippines.
[2] Chung Yuan Christian University, Chungli, Taiwan R.O.C.

* Corresponding author. Email: jvmlim@mymail.mapua.edu.ph
Manuscript submitted August 9, 2015; accepted December 24, 2015.
doi: 10.17706/ijcce.2016.5.5.331-340

Abstract: The study integrates a component correction algorithm on the electronic nose, intended for fruit ripeness determination, to counteract the tendency of the electrochemical metal oxide sensors to drift. The drift is any unwanted deviation of the chemoreceptor response from the true value leading to erroneous and inconsistent results. The study utilizes CCPCA — Component Correction Principal Component Analysis to approximate and remove the temperature drift component. In using this algorithm, the study theorizes that the drift acquired from the simulation of temperature drift without any sample (clean air) could generalize the drift pattern or structure in sampling Lacatan aroma samples. The PCA or Principal Component Analysis is utilized for 'dimensioanlity reduction' and 'data mining' that uncovers patterns unobserved on the multidimensional data structure. The cluster Silhouette is computed to quantitatively validate the intracluster cohesion and intercluster separation. Using PCA, the electronic nose distinguished 630 Lacatan aroma samples on the PCA Loadings — PC1 (94.65%) and PC2 (3%) and significantly improved the clustering upon removal of temperature drift. The cluster Silhouette is improved from 0.7903 to 0.8571 (unripe), 0.5358 to 0.6080 (ripe), and 0.7784 to 0.8357 (overripe).

Key words: Counteracting temperature drift, Lacatan banana, improved clustering, enhanced ripeness determination, silhouette index.

1. Introduction

The electronic nose or e–nose is one among todays emergent technology that mimics the human smell through an array of chemo–receptors and a pattern–recognition algorithm; this has been embraced in various industrial, agricultural, medical, and smart home functions. In comparison with its human counterpart, it has: 1) the epithelium that is comprised of the sensory neurons transducing and transmitting the chemicals in its immediate environment into patterns of elect rical signals to the brain; and 2) the brains pattern–recognition ability. Although numerous studies had delved into analysis of fruit volatiles with e–noses for ripeness determination: apple, bananas, blueberry, grape, peach, tomato, and mandarin [1]-[7]. This study intends at enhancing stability by building a temperature insensitive e–nose gearing towards the mechanisms portability.

Other than the fruits color, the aromas chemical structure vary with its ripeness; Zhang *et al*. enumerated this Volatile Organic Chemicals as: Esters, Alcohols, Aldehydes, Ketones, Apocarotenoids, Lactones, and

Tarpenoids [8]. Furthermore, is an enhanced synthesis of ethylene (C2H4) and enhanced respiration rate in ripening. Upon overripening and decaying, glucose is turned into alcohol and other foul–smelling odorants are introduced as enzymes and micro–organisms feed on the decaying organic matter. Other studies dealt with examining a distinct fruit features (color, internal composition, firmness) that is correlated to fruit ripeness; among these studies are: colorimetry, spectrophotometry, MRI (Magnetic Resonance Imaging), and Doppler vibrometry [9]-[12]. In general, an e-nose utilizes 1) an array of metal oxide chemoreceptors–transducers, 2) PCA for data – mining and dimensionality reduction, and 3) a neural network or other classification algorithm. Through recent studies pertaining to e–nose becoming involved in medical, agricultural, industrial, and smart home applications, the relevance of this technology is being realized.

A drift is the tendency of the metal oxide electro–chemical sensors to yield inconsistent readings due to 1) poisoning or aging of the sensor, and 2) change in environment humidity and temperature. In any system, inconsistency is not preferred as this may result to erroneous output and may demand periodic recalibration of the system to suit the current state or condition. Numerous mathematical algorithms had been developed to counteract the drift and one is a PCA based component correction algorithm CCPCA. In utilizing CCPCA for drift counteraction, this study assumes that the temperature drift approximated from sampling a clean air can be used to generalize the drift structure that is existent in sampling a fruit sample thus is deemed as the calibration group for CCPCA.

The intent of the study is counteract the temperature drift through a component correction algorithm based on PCA deemed as CCPCA. This study, in particular, aims at 1) building an electronic nose that captures the distinction of aromas from an unripe, ripe, and overripe Lacatan banana samples; 2) simulate the occurrence of the temperature drift; and 3) validate the clustering through the Silhouette indices.

The study's importance lies on the integration of drift–correction algorithm on the electronic nose, eliminating system's need for recurring calibration to suit the environment temperature; furthermore, the system homogenizes the "ripeness determination" through analysis of Volatiles emitted by a ripening fruit.

The study integrates metal oxide chemoreceptors: 10 MQ from the Winsen Electronics and 2 TGS from the Figaro Engineering Inc. though numerous forms of the sensor exist as 1) SAW — Surface Acoustic Wave, 2) QCM — Quartz Crystal Microbalance, 3) Optical, and 4) MEMS — Microelectromechanical Systems. The study centers on the Lacatan banana cultivar that is amongst the country's lead fruit exports and does not include the fruit handling prior the sampling. The drift, due to the temperature, is simulated from 32^{o}C to 38^{o}C and is removed via the component correction algorithm CCPCA.

2. Methodology

2.1. Electronic Nose Hardware

The chemical/s, that constitute the aroma of the banana sample, are transduced into their equal analog potential through an array of metal oxide chemoreceptors, then are transmitted to the analog IOs of the Arduino MEGA 2560 microcontroller. The electronic nose, built herein, utilizes 12 chemoreceptors, among those are: MQ2, MQ3, MQ4, MQ7, MQ9, MQ131, MQ135, MQ136, MA137, and MQ138 — from the Winsen Electronics; and TGS822 and TGS 826 — from Figaro Engineering Inc. The 12 chemoreceptors assume a potential divider configuration with a 20 kOhm trimmer potentiometer and is sourced by an external 9Vdc adaptor converted into 5V through an L7805 voltage regulator. The system, furthermore, has a DHT11 humidity and temperature sensor that is required in simulation and approximation of the drift. The modules of this electronic nose are condensed on a 1.1L volume of air–tight container; this is demonstrated in Fig. 1.

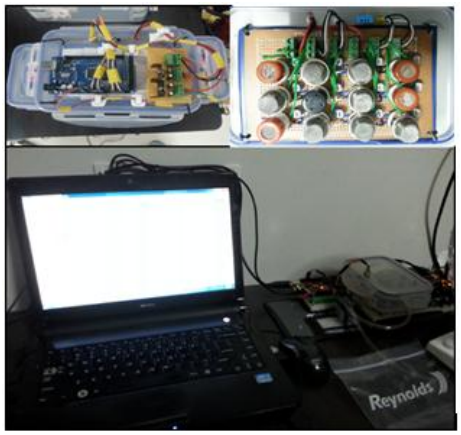


Fig. 1. Electronic nose hardware setup.

2.2. Principal Component Analysis

The PCA — principal component analysis is a multivariate data–mining and dimensionality–reduction tool that lies common among regression and clustering functions. The tool uncovers any pattern/s that is unseen or unobserved on a multidimensional space through the projection of the *n*–points onto k orthogonal features or principal components that span the direction of maximum variance. The PCA, done on an *m*–by–*n* matrix (*m* and *n* denote the number of samples and dimensions accordingly), *X*, implies the computation of the eigenvalues and eigenvectors of its *n*–dimensional square Covariance matrix. The eigenvalues quantify the amount of information spanned on the direction of the nth feature or principal component and is normally arranged in ascending order. The PCA involves the following:

1) Mean subtraction — centering the mean

$$\bar{X}_i = \frac{\sum_{j=1}^{m} X_{ji}}{m} \tag{1}$$

$$X_{\text{Mean\_Centered}} = X - \bar{X} \tag{2}$$

2) Computation of the covariance matrix

Let k and i in P_k and P_i denote the k^{th} and i^{th} dimension on the m–by–n matrix X. Variance and the n–dimensional square covariance matrix equates to (3), (4), and (5)

$$Var(P_k) = \frac{\sum_{j=1}^{m}(X_{jk} - \bar{X}_k)}{m-1} \tag{3}$$

$$Cov(P_k, P_i) = \frac{\sum_{j=1}^{m}(X_{jk} - \bar{X}_k)(X_{ji} - \bar{X}_i)}{m-1} \tag{4}$$

$$Cov_{\text{Matrix}} = \begin{bmatrix} Var(P_1) & \cdots & Cov(P_1, P_n) \\ \vdots & \ddots & \vdots \\ Cov(P_n, P_1) & \cdots & Var(P_n) \end{bmatrix} \tag{5}$$

3) Computation of the eigenvalues (λ) and eigenvectors (V)

4) Feature selection — principal components (P)

Numerous methods were implemented with PCA to determine '*What*' and '*How many*' of the *n* Eigenvectors (V) are loaded on Features vector (P). One is the *CPV* or the *Cumulative Percent Variance* that arranges both the eigenvectors (V) and eigenvalues (λ) in ascending order. In CPV, the 1st *I* features that

meet the required CPV, e.g. 90% or 95%, are loaded and the rest truncated on the features vector. CPV and the features vector are computed through (6) and (7).

$$CPV(I) = 100 \times \frac{\sum_{j=1}^{I} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \tag{6}$$

$$P = [V_1 \dots V_I] \tag{7}$$

5) PCA Scores (Y)

$$Y = PX_{\text{Mean\_Centered}} \tag{8}$$

2.3. Component Correction Principal Component Analysis

The component correction algorithm, introduced by Artursson et al., is a multivariate tool that eliminates unwanted occurrences of the chemoreceptor drift. The prime assumption of this algorithm lies on the thought that the exposure of the chemoreceptors to the same substance should result to uniform and consistent responses. This, furthermore, requires a reference group generalizing the drift structure or pattern in all other clusters. In using CCPCA, this study assumes that the temperature drift approximated from stimulating drift occurrence on sampling a clean air is also existent upon sampling banana samples. Thus is deemed as the reference cluster in the drift approximation and removal. CCPCA is given by (9) where V_{Cal} pertains to the eigenvectors (V) of the *calibration* group — the cluster that generalizes the drift pattern or structure.

$$X_{\text{Corrected}} = X - (X \times V_{Cal}) \times V_{Cal}^{T} \tag{9}$$

2.4. Silhouette Index

A Silhouette is a measure of intra–cluster cohesion and inter–cluster separation that computes the Squared Euclidean distances of the k^{th} point to the others and designates an equivalent Silhouette Index. An index near (+)1 implies that the k^{th} point belongs to the correct cluster. If k is index of the kth point on the PCA bi–plot, let $a(k)$ be the mean dissimilarity of k from all other points of the same cluster and $b(k)$ the mean dissimilarity of k from all other points on the neighboring cluster. The Silhouette index of the k^{th} point is computed through the (10).

$$s(k) = \frac{b(k) - a(k)}{\max\{b(k), a(k)\}} \tag{10}$$

Table 1. Interpretation of the Silhouette Index of the k^{th} Sample

$s(k)$	Interpretation
$s(k)$ close to 1	The kth sample lies within assigned cluster
$s(k)$ is near 0	The kth sample lies on the border of 2 adjacent clusters
$s(k) < 0$	The kth sample lies within neighboring clusters

Table 2. Interpretation of the Mean Silhouette Coefficient of the Cluster

Silhouette Coefficient	Interpretation
0.71 – 1.0	A strong structure is found
0.51 – 0.7	A reasonable structure is found
0.26 – 0.50	Structure may be artificial
< 0.26	No substantial structure found

2.5. Experimentation

All Lacatan aroma sampling are done within a cold environment (air–conditioned) to attain a minimum temperature drop of 28°C, the sampling extends from 20 to 25 minutes until the temperature read by the DHT11 equates to 38°C. Lacatan banana sample is then introduced at 28oC and accumulates within the chamber for an extent of 5 minutes. At this instant, the temperature read by the DHT11 amounts to 32°C. The fruit aroma sampling begins at 32°C, the point the aroma is assumed to have saturated the entire volume of the container and ends at 38°C. A Matlab function extracts and returns the chemoreceptor responses at the temperatures 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, and 38°C. Prior sampling, all 12 MQ and TGS are subjected to the environment air for an extent of 5 minutes to guarantee the full recovery of the chemoreceptors. This, furthermore, promotes consistency in the sampling of 90 Lacatan banana samples. The aroma sampling begins at 32°C since it is learnt that at this point, the aroma had saturated the entire volume demonstrated by the saturation and the continuous, linear, yet gradual build-up in the chemoreceptor responses. This continuous linear build–up is then accounted to the temperature drift.

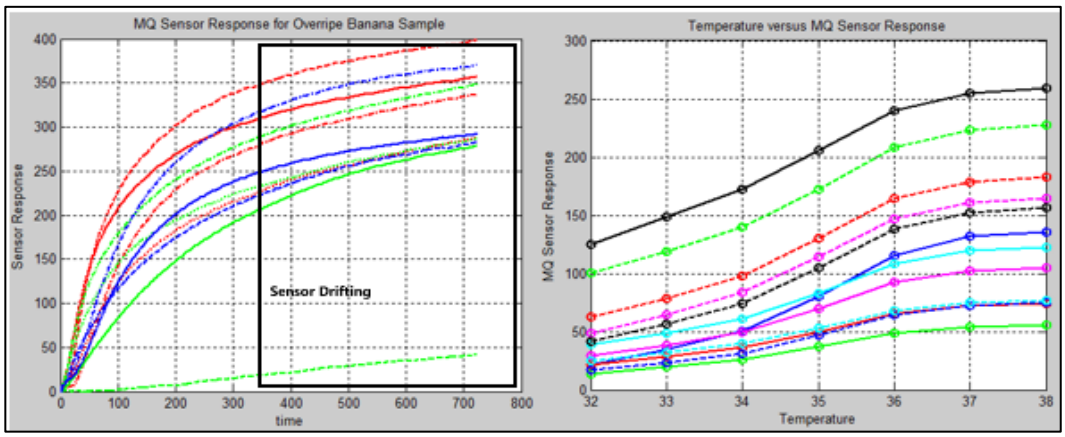


Fig. 2. (a) The temperature drift in chemoreceptor readings, and (b) the temperature — MQ/TGS sensor plot.

A total of ninety (90) Lacatan banana samples are involved in the experiment; of the ninety (90), a third or thirty (30) are brought into the chamber unripe, another thirty (30) are sampled ripe, and the remaining thirty (30) overripe. The distinction between an unripe, a ripe, and an overripe Lacatan banana samples is observed in Fig. 3.

(a) (b) (c)

Fig. 3. (a) Unripe, (b) Ripe, and (c) Overripe Lacatan banana samples.

3. Results and Discussion

3.1. Evaluating e–Nose Function for Fruit Ripeness Determination

Prior the simulation, approximation, and removal of the unwanted temperature drift, it must be guaranteed that the array of MQ and TGS chemoreceptors could capture the distinction of aroma samples from an unripe, ripe, and overripe Lacatan bananas. The Lacatan aroma smell–print is acquired from the

extraction of the chemoreceptor responses at 32°C, the point the aroma is assumed to have saturated the container. The matrix of Lacatan aroma is comprising 90 rows — number of Lacatan banana samples and 12 columns — number of MQ and TGS chemoreceptors. Through PCA, the matrix of Lacatan samples is plotted on 2 features or components — PC1 (94%) and PC2 (3%), and the distinction of the aroma is underlined on direction of the PC1 as demonstrated in Fig. 4.

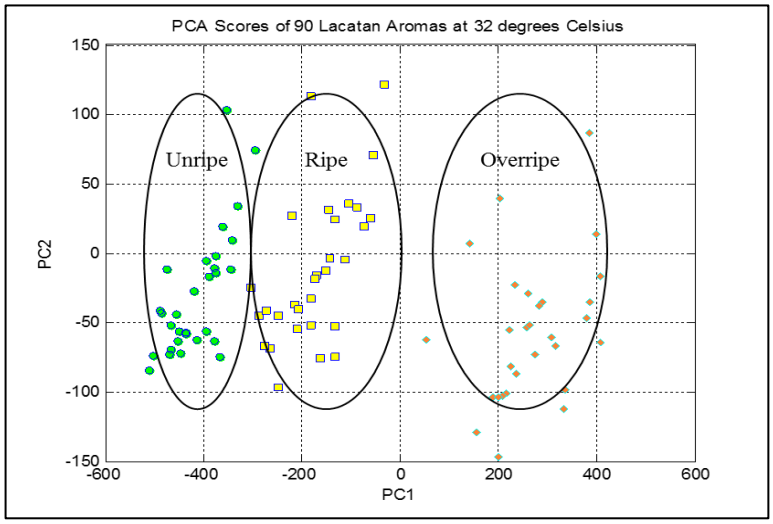


Fig. 4. PCA bi–plot (Scores) of Lacatan aroma scores acquired at 32 degrees Celsius sampling temperature.

The silhouette indices of the 90 Lacatan aroma scores and the mean silhouette index of the 3 distinct fruit clusters are computed to evaluate the clustering. The computed indices and mean index are interpreted via Table 1 and Table 2. The unripe cluster has a mean silhouette index of 0.8156, ripe cluster has 0.6762, and overripe 0.8822. The silhouette plot is demonstrated on Fig. 5.

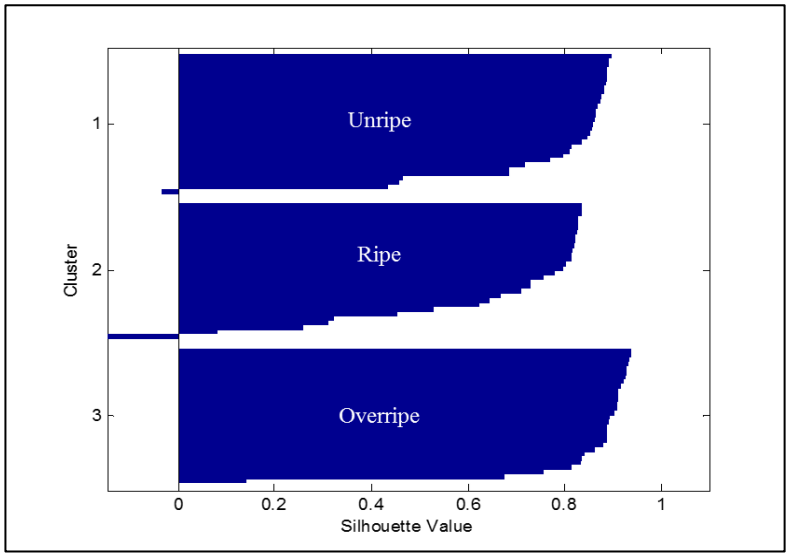


Fig. 5. Silhouette of the 90 Lacatan aroma scores acquired at 32 degree Celsius sampling temperature.

3.2. Simulation, Approximation, and Removal of Temperature Drift

The drift, accounted to inconsistency in temperature, involves the sampling of the Lacatan aroma as the instantaneous temperature read within the chamber is gradually altered from 32°C to 38°C. It is apparent

from Fig. 2. that the MQ/TGS chemoreceptor values don't converge to explicit values and continues to build–up with temperature. The direction and approximate magnitude of the same temperature drift could be exhibited on a 2 dimensional plane: PC1 (94.65%) and PC2 (3%) through PCA. The matrix of Lacatan aromas, with temperature drift, is 630–by–12 in dimension. The scores of 630 Lacatan aromas on the PCA plane is demonstrated on Fig. 6 and Fig. 7. The temperature drift's direction is superimposed on the plot.

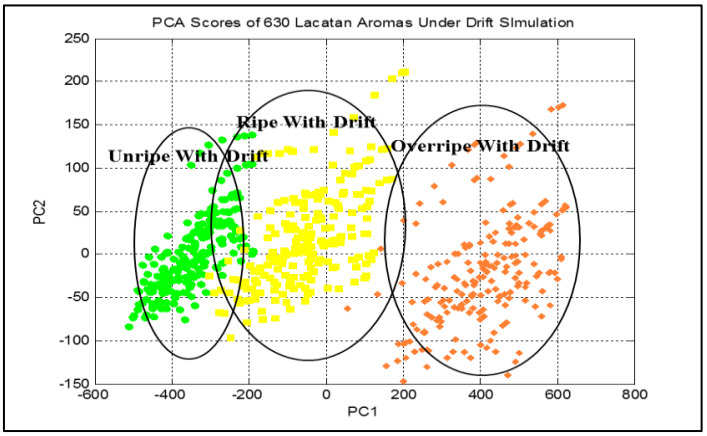
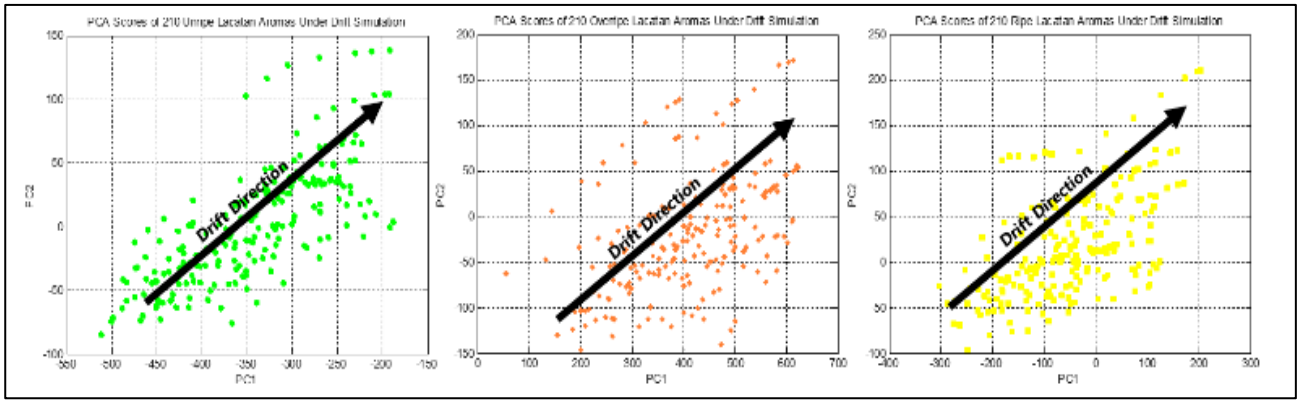


Fig. 6. PCA bi–plot (Scores) of 630 Lacatan aromas under drift simulation.



(a) (b) (c)

Fig. 7. PCA bi–plot (Scores) of (a) unripe, (b) ripe, and (c) overripe Lacatan aromas with drift direction superimposed on the plot.

The Lacatan aroma scores (on Fig. 6), bounded by the -300th and -200th index on PC1, do not belong to distinct fruit aroma cluster and may be mixed–up to unripe or ripe cluster. Thus, it is conclusive that occurrences of the temperature drift results to poor clustering on the PCA features — PC1 and PC2. Quantitatively, through the scores silhouette, the clustering of the 630 points evaluated and interpreted through Table 1 and Table 2. This Scores silhouette, furthermore, becomes the benchmark that is to be compared upon removal of drift component. The Unripe cluster has 0.7903 silhouette index, Ripe has 0.5358, and Overripe has 0.7784. The measure of cluster cohesion, demonstrated by the silhouette plot, and the mean silhouette indices of the 3 distinct fruit clusters, is demonstrated in Fig. 8.

Component correction–removal of unwanted temperature drift, is done by sampling clean environment air from 32^{o}C to 38^{o}C and their projection on 2 orthogonal features of maximum variance. Since from the 32nd to the 38th sampling temperature, the clean air continuously flow underneath the chemoreceptors, it is presumed that the amount of variance spanned on both PC1 and PC2 is accounted to the temperature drift. Thus in the drift removal, the VCal is a column vector containing the eigenvectors PC1 and PC2. In using

CCPCA for drift removal, this study assumes that the drift approximated in the drift simulation of this clean air is also existent upon the sampling of unripe, ripe, and overripe Lacatan banana samples and thus could generalize the drift pattern in the 3 clusters. The uncorrected, with the drift, and the corrected, without the drift, PCA Scores of Lacatan aromas are demonstrated on Fig. 9.

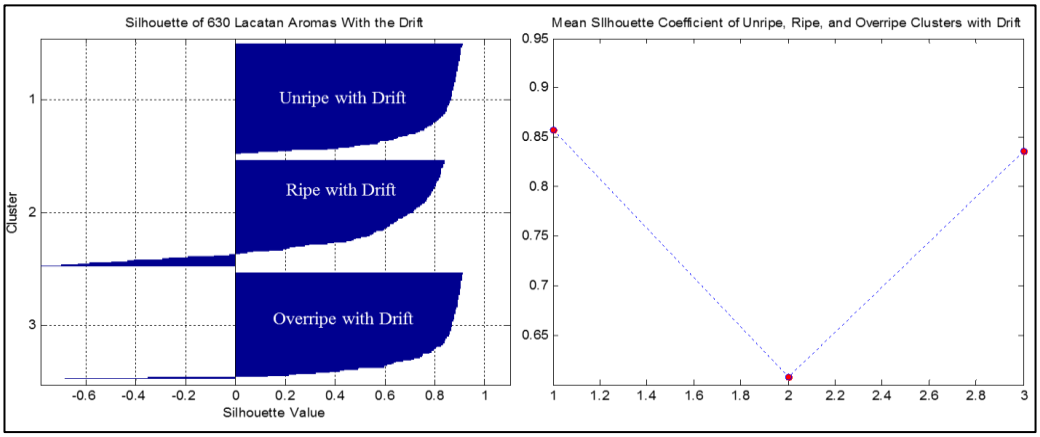


Fig. 8. The (a) clusters' Silhouette and (b) mean Silhouette indices of Lacatan aroma scores under drift simulation.

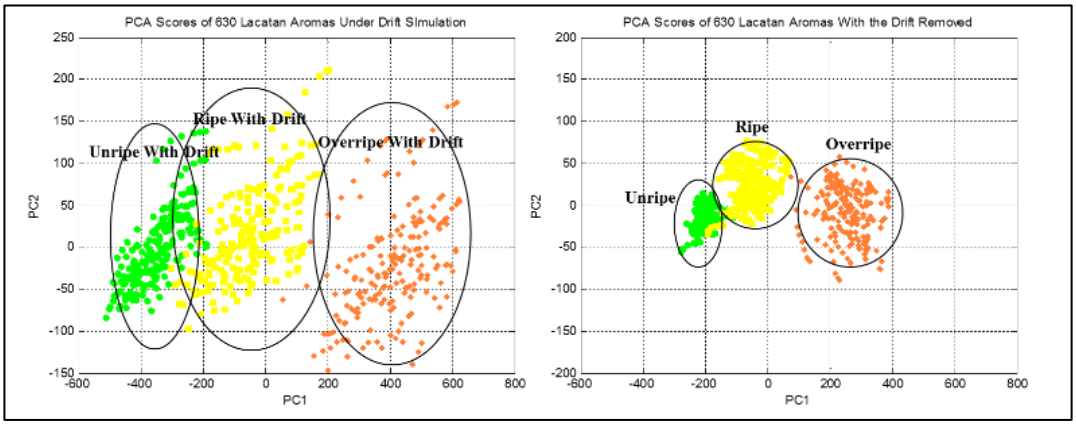


Fig. 9. PCA bi–plot (Scores) of Lacatan aromas (a) with the temperature drift prior CCPCA and (b) with the removal of temperature drift.

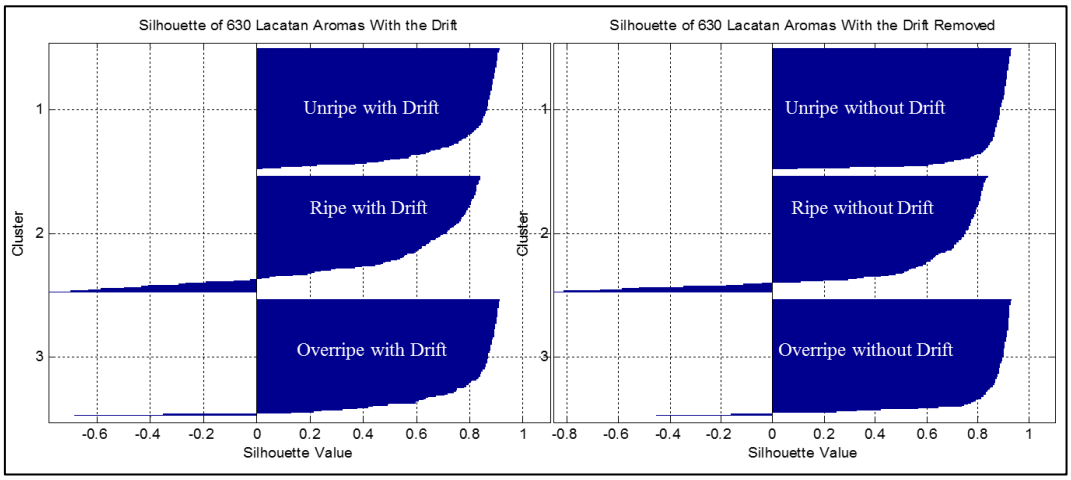


Fig. 10. Silhouette of unripe, ripe, and overripe clusters (a) with the temperature drift and (b) without temperature drift.

Apparently, on Fig. 9, the bi–plot of the Lacatan aroma scores with the temperature drift removed is more condensed than the bi–plot with the unwanted drift component. This is indicated on the scores silhouette as the 1st cluster (unripe) is improved from 0.7903 to 0.8571, the 2nd cluster (ripe) from 0.5358 to 0.6080, and the 3rd (overripe) from 0.7784 to 0.8357. The enhancement brought by removal of *drift* can be observed by a side-by-side comparison of the silhouette and mean silhouette indices of measurements taken prior and after CCPCA, this is demonstrated in Fig. 10 (Silhouette plot) and Fig. 11 (Mean silhouette indices).

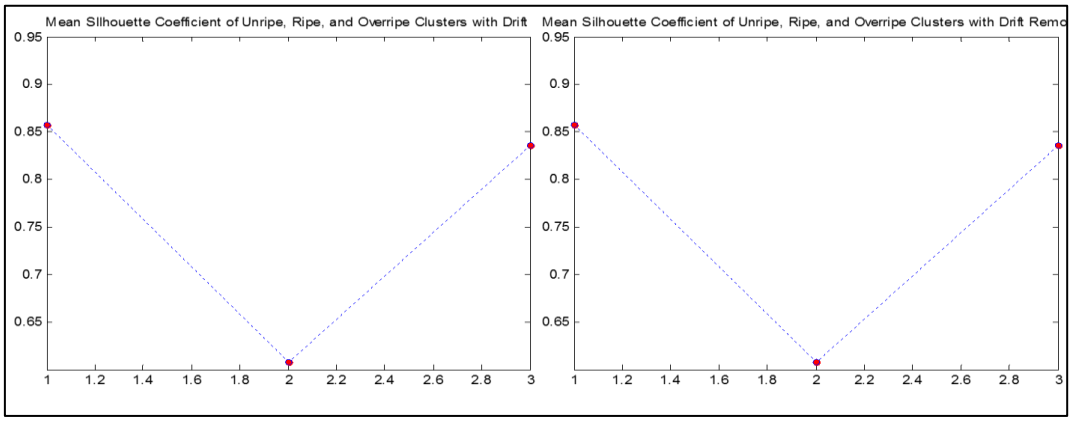


Fig. 11. Mean Silhouette indices of unripe, ripe, and overripe clusters (a) with the temperature drift and (b) without temperature drift.

4. Conclusion

An electrochemical metal oxide chemoreceptor drifts upon exposure of the system to distinct range of temperature values resulting to recurring need for recalibration to suit the current environment condition. The system, along with the 10 MQ and 2 TGS chemoreceptors, could distinguish an unripe, ripe, and overripe Lacatan aroma samples through their projection on the 2 eigenvectors through PCA–PC1 (94%) and PC2 (3%) with the cluster Silhouettes of 0.8156, 0.6762, and 0.8822 accordingly. Introducing the drift through a gradual increase in sampling temperature from 32°C to 38°C influence the Lacatan aroma scores on the principal components–PC1 (94.65%) and PC2 (3%) resulting to a reduced value of cluster Silhouettes of 0.7903, 0.5358, and 0.7784 accordingly. In removing temperature drift through CCPCA, the system used the drift acquired in the drift simulation for clean air from 32°C to 38°C as the 'calibration' group the generalizes the drift pattern in sampling Lacatan aroma scores. The effectivity of this method is quantified by the improvement of the cluster Silhouettes from 0.7903 to 0.8571 (unripe), 0.5358 to 0.6080 (ripe), and 0.7784 to 0.8357 (overripe).

Acknowledgment

The authors recognize the significant contributions of all the people and government organization who had been involved in the completion of this study. The authors thank the DOST–ERDT (Department of Science and Technology's Engineering Research and Development for Technology) for the funding of this research.

References

[1] Brezmes, J., Llobet, E., Vilanova, X., Orts, J., Saiz, G., & Correig, X. (2001). Correlation between electronic nose signals and fruit quality indicators on shelf-life measurements with pink Lady apples. *Sensors and Actuators B: Chemical, 80(1)*, 41-50.

[2] Llobet, E., Hines, E. L., Gardner, J. W., & Franco, S. (1999). Non-destructive banana ripeness determination using a neural network-based electronic nose. *Measurements Sci. Technol.*, *10(6)*, 538-548.
[3] Simon, J. E., Hetzroni, A., Bordelon, B., Miles, G. E., & Charles, D. J. (1996). Electronic sensing of aromatic volatiles for quality sorting of blueberries. *J. Food Sci., 61(5)*, 967-969.
[4] Patterson, T. (2007). The electronic nose knows: handheld technology measures grape maturity. *Wines and Vines, 88(12)*, 46.
[5] Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., & Mignani, E. (2008). Electronic nose as a non-destructive tool to characterize peach cultivars and to monitor their ripening stage during shelf-life. *Postharv. Biol. and Tech., 47(2)*, 181-188.
[6] Gomez, A. H., Hu, G. X., Wang, J., & Pereira, A. G. (2006). Evaluation of tomato maturity by electronic nose. *Computers and Electronics in Agriculture*, *54(1)*, 44-52.
[7] Gomez, A. H., Wang, J., & Pereira, A. G. (2007). Mandarin ripeness monitoring and quality attribute evaluation using an electronic nose technique. *Trans. ASABE, 50(6)*, 2137-2142.
[8] Zhang, F., *et al*. (2013). Advances in fruit aroma volatile research. *Molecules*, *18*, 8200-8229.
[9] Malevski, Y., Gomez-Brito, L., Peleg, M., & Silberg, M. (1977). External color as maturity index of mango. *J. Food Sci., 42*, 1316-1318.
[10] Jha, S. N., Kingsly, A. R. P., & Chopra, S. (2006). Non-destructive determination of firmness and yellowness of mango during growth and storage using visual spectroscopy. *Biosystems Engineering*, *94(3)*, 397-402.
[11] Chen, P., McCarthy, M. J., & Kauten, R. (1989). NMR for internal quality evaluation of fruits and vegetables. *Trans. ASAE.*, *32*, 1747–1753.
[12] Santulli, C., & Jeronimidis, G. (2006). Development of a method for nondestructive testing of fruits using scanning laser virometry (SLV). From http://www.ndt.net/article/v11n10/santulli.pdf

John Victor M. Lim was born in October 1992 in Manila, Philippines. He took up a bachelor degree of science in computer engineering (BS CpE) majored in microelectronics at the Mapua Institute of Technology, Manila, Philippines. He also took master of science in computer engineering (MS CpE) as a joint BS–MS program offered in the mentioned technological institution under the Department of Science and Technology — Engineering Research and Development for Technology scholarship program. He completed both degrees in March 2015.

Noel B. Linsangan is the present program chair of the Computer Engineering (CpE) program at the Mapúa Institute of Technology, Manila, Philippines. He obtained his baccalaureate degree in computer engineering from the same school in 1988 and his graduate degree, master of engineering in computer engineering from the Pamantasan ng Lunsodng Maynila (University of the City of Manila) in 2000. Right after graduation from college, he joined the Faculty of Computer Engineering Program and in 2000, he was appointed as the program chair. He is part of the team that pioneered the implementation of OBE in Mapúa and also one of the point persons of the institute when its EE, ECE, and CpE programs were preparing for ABET accreditation.