

Concentration Analysis of *Eurycoma Longifolia* Using Principle Component Based Artificial Taste Sensor

Mohd. Zaid Abdullah¹, Mohd. Noor Ahmad², Ali Yeon Md. Shakaff³, Zhari Ismail⁴

¹School of Electrical and Electronic Engineering
Universiti Sains Malaysia, Engineering Campus
14300 Nibong Tebal, Seberang Perai Selatan, Pulau Pinang, Malaysia
Tel: +604-5937788, Fax: +604-5941023, E-mail: mza@eng.usm.my

²School of Chemical Sciences
Universiti Sains Malaysia
18000 Minden, Pulau Pinang, Malaysia
Tel: +604-6577888, Fax: +604-5941023

³School of Engineering
Kolej Kejuruteraan Utara Malaysia
Taman JKKK, Kubang Gajah 02600 Arau Perlis, Malaysia
Tel: +604-9781300, Fax: +604-9778011

⁴School of Pharmaceutical Science
Universiti Sains Malaysia
18000 Minden, Pulau Pinang, Malaysia
Tel: +604-6577888, Fax: +604-5941023

Abstract

Water extracts of *Eurycoma Longifolia* (Tongkat Ali) were analysed by artificial-taste sensor employing 4 mm diameter and 20 μm thickness lipid membrane technology. The sensing element which comprises of 8 sensing channels and 1 reference channel was attached to high impedance differential voltmeter forming the data acquisition and capturing system. Measurements were performed on extracts of four concentration levels: 0.001%, 0.025%, 0.05% and 0.075%. Altogether 384 measurements were performed comprising of 24 potentials per concentration level. These data were treated with direct and principal component based discriminant analyses and results compared. It was discovered that direct application of discriminant analysis produced an overall accuracy of 22.9% compared to 92.7% of principle component based discriminant analysis. Results obtained confirmed our prediction that the data set produced by the multichannel taste sensing system were highly correlated, and selecting the best subset containing variables with highest discrimination powers could significantly improved the separation between groups.

Keywords:

Taste sensing; Discriminant Analysis; Principal Component Analysis.

Introduction

The nutritional and commercial importance of *Eurycoma Longifolia* have long been recognised, particularly its aphrodisiac property. To-date this material forms one of the major constituents in many health-based agricultural and food products. Determining and assessing the authenticity of the *Eurycoma Longifolia* based foodstuffs has been studied extensively by conventional chromatographic methods[1]. These methods are very accurate but also costly and time consuming, and they require the use of toxic, flammable pollutants reagents and gases. This has prompted scientists and engineers into designing instrumentation capable of broad-band chemical detection and quantification. One such example is the development of electronic or artificial-tongue[2-3]. In this system, arrays of cross-reactive chemical sensors are coupled to pattern recognition programs, paralleling the biological gustatory system, in which semi selective gustatory receptors are combined with higher order or neural processing. Recent studies have demonstrated the potential of qualitative and quantitative analysis based on artificial-tongue spectra to discriminate between mineral water from different manufacturers[4-5]. Most of these studies were based on the application of neural-network on the complete data information. This method is rather complex and require extensive computing facilities and complicated interfacing hardware. However, Abdullah *et. al.* [6] followed approach based on multivariate analysis for solving pattern recognition problems. Instead of using all available data, they have selected best variables for

classification using the Wilk's Lambda analysis. Reducing the dimensionality of data is preferred in this case since it minimises errors due to redundancy and interdependence among variables. They concluded that discriminant analysis after Wilk's Lambda analysis is more precise in classification than direct application of discrimination analysis. The same procedures can be applied to electronically solve pattern classification problems in artificial taste sensor. Here, dimensionality reduction was achieved via principal component analysis which performed linear combinations of data instead of deleting a given variable or subset of variables.

Approach and Methods

Electronic taste sensing system

The data acquisition system for electronic taste sensor used in this study is similar to the one reported in [4]. Therefore, interested readers are referred to this paper for further technical details. Here, a summarised version of this system is presented in order to facilitate discussion on algorithm for pattern recognition. The block diagram of the taste sensing system is shown in Figure 1.

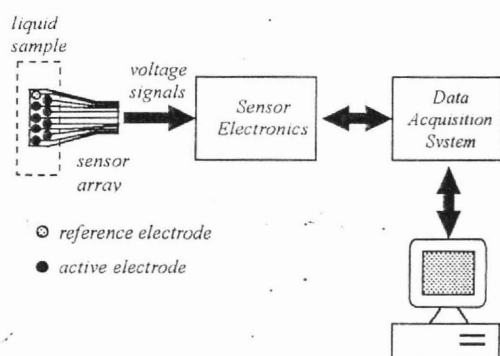


Figure 1 Schematic block diagram of electronic taste sensing instrument

The electrodes depicted in Figure 1 are disposable screen-printed lipid-membrane taste sensors. The sensor array comprises of 8 active electrodes and a reference electrode. The material used for active electrodes depends largely on sensor applications. Here we used lipid materials similar to the ones reported by Toko[2]. Meanwhile the Ag/AgCl was chosen material for reference electrode. The electrodes are fabricated in such a way that they make electrical contact with the liquid inside the test beaker but do not effect its normal mass transfer. The sensor array is connected to sensor electronics by short lengths of driven shield co-axial cable to reduce the effect of extraneous environmental noise and interference. Electrochemical reactions which occur at the liquid-electrode interfaces induce dc voltages in mV range. The strength of these voltages depend largely on the magnitude of the electrochemical reactions, which in turn, depend on

electrode materials. Since each electrode is fabricated from different lipid materials, hence, the induced voltages are also different. The combined signal voltages of all the eight electrode yield a signature or finger-print of the test sample.

The voltage measurement circuitry extends from the voltage buffers in sensor electronics module to PC-based data acquisition system. In DAS the output sensor electronics is firstly multiplexed, and secondly, routed into differential input amplifier, which amplifies the potential difference between the voltage and reference signals. The important attribute of such an amplifier is its ability to reject common-mode signals such as extraneous electrical noise appearing on both input lines. The common-mode rejection ratio (CMRR) in the present system is approximately 90dB at dc – over 30 dB higher than most typical high speed wide-band differential amplifiers. This figure was accomplished in our circuit by switching the inputs to the amplifier to negate the effects of input impedance imbalance between two lines. The output of differential amplifier is fed into a programmable gain amplifier and finally to general purpose analogue-to-digital converter (ADC) where it is compared to previously stored 'standard' value.

Figure 2 shows typical voltage signatures captured by the system in response to 4 *Eurycoma Longifolia* concentration levels– 0.01%, 0.03%, 0.05% and 0.08%.

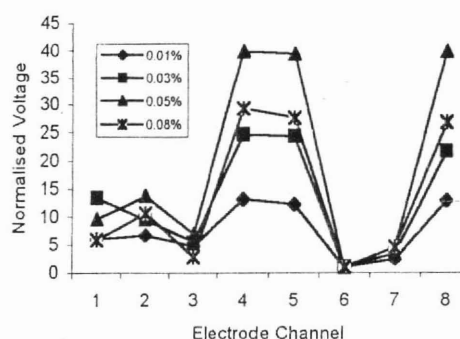


Figure 2 Steady state response of electronic taste equipment to four *Eurycoma Longifolia* concentrations.

From Figure 2, channels 4,5 and 8 registered consistently higher magnitudes in the following concentration sequence: 0.05%, 0.08%, 0.03% and 0.01%. Theoretically, the method of direct thresholding can be used to determine the different concentration levels of *Eurycoma Longifolia*. However, a close inspection of Figure 2 reveals that voltages registered by channel 1-3 and 6-7 for different concentration levels overlap with each other. Clearly, no single threshold value exists which can uniquely separate the *Eurycoma Longifolia* based on responses of sensor array. Furthermore, the presence of voltage noise can affect the responses of the sensor array, resulting in a great number of *Eurycoma Longifolia* being misclassified. Clearly a different approach is needed to solve this pattern classification problem.

One convenient method to solve this problem was to treat the voltage profiles as features and apply multivariate discriminant analysis to directly obtain a criterion from training multiple samples. The advantage of using this method was that it does not need to consider voltage profile details. However, its main drawback was that the discarding of variables which were correlated highly with those retained did not necessarily guarantee that the deleted variables were, in fact, redundant in the classification model. Similarly, retaining the variables irrespective of their correlation characteristics did not necessarily yield a subset with maximum rate of correct classification. We demonstrated in this paper that the use of principle component analysis before discriminant analysis helped overcome this problem. The details are given below.

Principle component and discriminant analyses

Primarily principle component analysis (PCA) deals with dimension reduction while discriminant analysis (DA) deals with separating distinct sets of observations and subsequently allocating new observations to the defined groups. Dimension reduction is achieved by linear combinations of the data instead of data deletion or removal as in direct or step-wise DA. In this way, PCA preserves the original data set before and after classifications. Theoretically, both PCA and DA are established statistical techniques and interested readers are referred to a text book by Dillon and Goldstein (1984) for technical details. A summarised version is presented here, focussing on taste sensing application.

In deriving the principle components, it is desirable to firstly investigate the interdependence of the variables among themselves in a multidimensional space. In so doing let $V(1,2,3,\dots,n)$, $\Sigma(1,2,3,\dots,n)$ and $P(1,2,3,\dots,n)$ be the original data (voltages), variance-covariance and correlation matrices based on m measurements and n sensors. In this case, P is given as follows,

$$p_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad (1)$$

where σ is the ij th variance. Note that P is symmetrical matrix having all diagonal elements equal to one. Hence, the interdependence between i th and j th variables is the highest when p_{ij} is maximum. Figure 3 shows typical relationships among variables of the 8-channel taste sensing equipment. Clearly, there is a strong relationship between channel 2 and channel 6, implying that the voltages generated by these electrodes were highly correlated. Channel 3 and channel 7 also appeared visually correlated. Hence, it can be seen that there were at most 6 uncorrelated channels from 8 available channels. Secondly, the eigen values of P is calculated in order to examine the percentage variation among the available channels. Mathematically,

$$|P - \lambda I| = 0 \quad (2)$$

where I is the identity matrix and λ is the eigen vector. Generally, the number of eigen values greater than 1.0, a common statistical cutoff point in PCA[5], determines the number of principal components. The variation of λ is demonstrated in the SCREE plot shown in Figure 4. Clearly from this graph that there were 2 principal components corresponding to eigen values greater than 1.0. These 2 principal component constituted 91% of the total sample variation. Thus the dimensionality of the data was reduced from 8 correlated voltages to 2 principal components with 75% loss of variation. Furthermore, it was observed that the proportion of variance explained by any principle component was not greater than 10% and the individual voltage loadings on the principle components were relatively uniform. In other words, this procedure gives a conservative estimate of the downward bias in λ due to combination as compared to all possible subsets procedure.

Finally, the principal component vector y is computed as follows:

$$y_{rj} = \frac{(b_j)^T (v_r - \bar{v})}{\sqrt{\sum_{ii}}} \quad (3)$$

where b is the j th eigen vector calculated as,

$$Pb = \lambda b \quad (4)$$

and for $j = 1, 2, 3, \dots, n$ and $r = 1, 2, 3, \dots, N$.

Once all principle components have been calculated using (3) and (4), classification can be performed using DA. The basic idea of DA is to generate the classification function for each group corresponding to concentration level of the analyte. For a population comprising of g groups, the i th classification function containing r principal components can be expressed as follows:

$$x_i = a_i + \sum_{j=1}^r C_{ij} y_j \quad \text{for } i = 1, 2, 3, \dots, r \quad (5)$$

where a_i is the i th constant, y_j is the j th principal component and C_{ij} is the i th classification function for the j th principal component. Classification coefficients can be computed by first forming the pooled variance-covariance matrix, W , and second, applying the following matrix multiplication:

$$C_i = W^{-1} \bar{v}_i \quad (6)$$

where \bar{v}_i is the column vector containing average measured voltages for i th class. Similarly, the classification constants can be calculated using the standard procedure as follows:

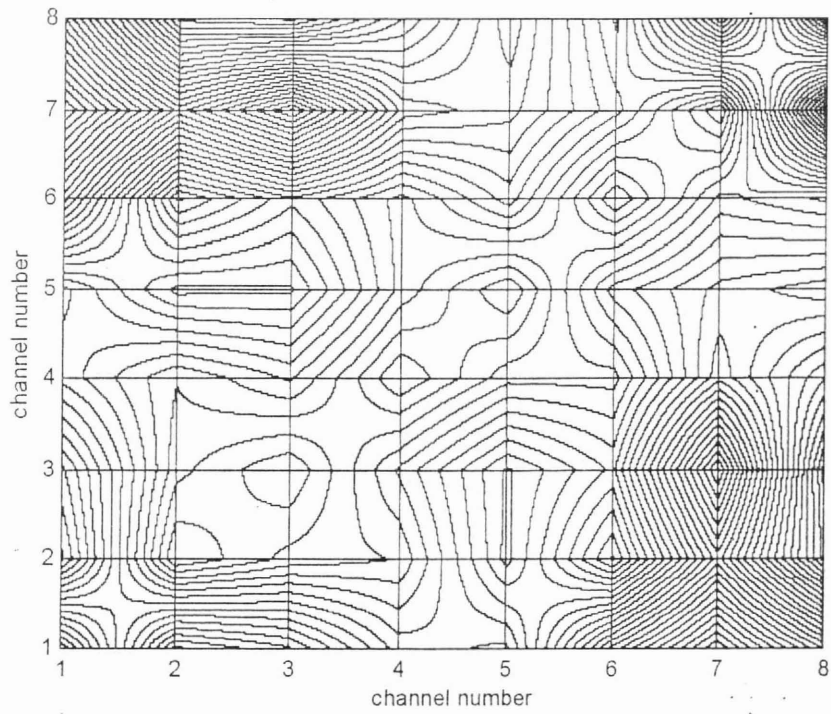


Figure 3 Correlation distribution of the taste sensing equipment showing the relationships among the 8 available channels.

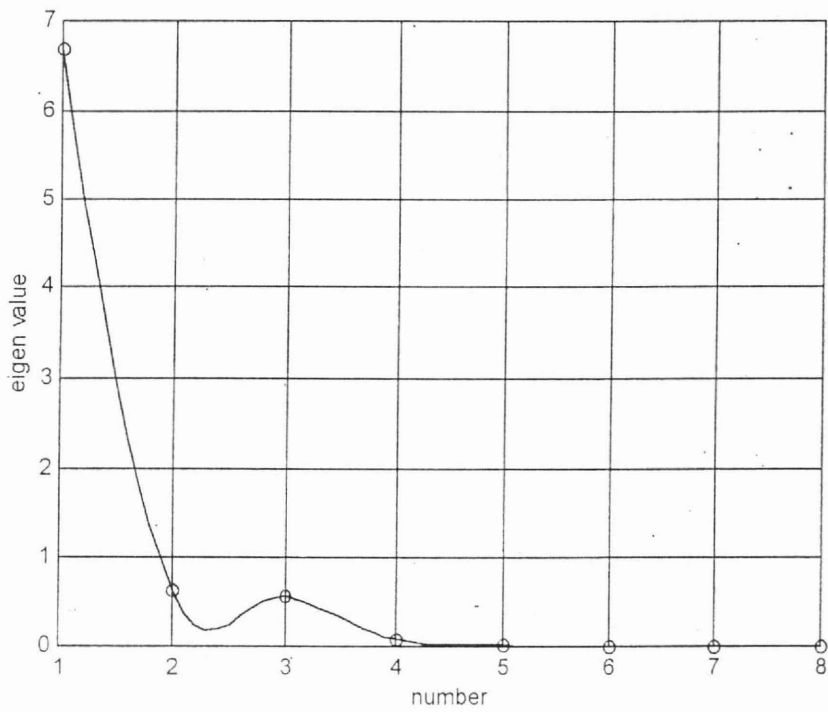


Figure 4 Variation of λ with the number of principal component. Note that the eigen values converge sharply to zero as the number of principal component approaches 2.

It is well known that DA returns optimal classification when the within-group covariance matrices are equal or homogeneous. In actual practice, we discovered that this condition is difficult to fulfill due to systematic, random and inherent measurement errors. We observed that the voltages recorded by the sensor drift as much as 5% to 12% with time. However, we also discovered that as long as the homogeneity is not significantly substantial then the robustness of the algorithm would not be severely affected. In fact heterogeneity improves the performance of the algorithm because of a slight bias [7].

Materials and Methods

Sample Preparation

Commercial types of *Eurycoma Longifolia* were purchased from local suppliers and stored at 5°C. The water-soluble fraction was extracted using standard spray-drying technique. The resulting powder was mixed with distilled water of known volume to produce the desired concentration. For instance, diluting x g of *Eurycoma Longifolia* powder in 100 ml distilled water produces x % of analyte concentration. This formula was used to prepare *Eurycoma Longifolia* of four concentration levels – 0.01%, 0.03%, 0.05% and 0.08%.

Measurement strategy

The measurement strategy taken was to employ a new sensor array for each concentration level. Measured voltages were grouped into two sets: Set A and Set B. Each set was further divided in two subsets referred here as Subset A1, Subset A2, Subset B1 and Subset B2. Each subset comprised of 96 voltages corresponding to 24 voltages for each concentration level. Therefore, altogether 384 measurements were performed requiring 8 sensor arrays. Set A was used for direct DA and Set B was reserved for PCA. Meanwhile, the first and second subset for each set was used for training and testing respectively. Classification experiments were firstly conducted using training set and secondly using testing set. In this way, the system accuracy could independently be investigated. For both direct DA and PCA, the discriminant program was trained using 24 concentration samples for each group.

Results and discussion

The *Eurycoma Longifolia* training and independent sample test results via direct DA are presented in Figure 5. For the training process, all 96 samples were correctly classified. However, for the independent test samples, 22 out of 24 samples of 0.01% concentration level were correctly classified. All the other samples were wrongly classified. Hence the correct classification rate of independent test samples was approximately 23% compared to 100% of the training samples. This indicates that the good separation produced by the training set is artificial since the rate of correct classification is biased against the best subset

discriminant analysis. The system seems to memorise the input signatures instead of generalising them. Clearly, the inclusion of all variables in the data set do not necessarily contribute to separation of the groups.

Concentration		Predicted			
		0.01%	0.03%	0.05%	0.08%
Actual	0.01%	24	0	0	0
	0.03%	0	24	0	0
	0.05%	0	0	24	0
	0.08%	0	0	0	24

(a)

Concentration		Predicted			
		0.01%	0.03%	0.05%	0.08%
Actual	0.01%	22	2	0	0
	0.03%	0	0	24	0
	0.05%	0	0	0	24
	0.08%	0	0	24	0

(b)

Figure 5 Classification results of direct DA of (a) training test samples and (b) independent test samples.

Results of discriminant analysis after PCA are tabulated in Figure 6.

Concentration		Predicted			
		0.01%	0.03%	0.05%	0.08%
Actual	0.01%	24	0	0	0
	0.03%	0	24	0	0
	0.05%	0	0	24	0
	0.08%	0	0	0	24

(a)

Concentration		Predicted			
		0.01%	0.03%	0.05%	0.08%
Actual	0.01%	24	0	0	0
	0.03%	0	24	0	0
	0.05%	0	0	24	0
	0.08%	0	0	16	8

(b)

Figure 6 Classification results of DA after PCA of (a) training test samples and (b) independent test samples.

Similarly, discriminant analysis after PCA of training samples resulted in 100% correct classification. However, for independent test samples, all 24 samples belonging to 0.01%, 0.03% and 0.05% concentration levels were correctly classified. Only 8 out of 24 samples of the 0.08% concentration level were wrongly classified. Hence, this method yielded an overall percent correct classification of more than 92%. The calculations using individual group covariances and pooled covariance showed a very closed result. Hence, it can be concluded that discriminant analysis after principal component analysis is more precise in classification than the direct application of discriminant analysis.

References

- [1] Chan, K.L. 1998. High performance liquid chromatography in physicochemical analysis of *Eurocoma Longifolia*. *Planta Medica* 64:741-745.
- [2] Toko, K. 1998. Electronic tongue. *Biosensor & Bioelectronics* 13: 701-709.
- [3] Toko, K. 2000. Taste Sensor. *Sensors and Actuators B* 64: 205-215.
- [4] Rahman, A.S.A.; Shakaff, A.Y.M.; Ahmad, M.N.; Teo, J.S.; Hitam, M.S. and Ismail, Z. 2002. A neural network-based electronic taste sensing systems for the verification of mineral water. In *Proceedings of the International Conference on Artificial Intelligence in Engineering and Technology*, (Sazali Yaacob, R. Nagarajan and Ali Chekina, eds.) pp. 757-763.
- [5] Teo, J.S.; Ahmad, M.N.; Hitam, M.S. and Shakaff, A.Y.M. 2001. Classification of taste using a neural network: a case study in mineral water and drinking classification. In *Advances in Scientific Computing, Computational Intelligence and Applications*, (N. Mastorakis, V. Mladenov, B. Sutter and L.J. Wong, eds.) pp. 459-464, WSES Press.
- [6] Abdullah, M.Z.; Guan, L.C. and Mohd Azemi, B.M.N. 2001. Stepwise discriminant analysis for colour grading of oil palm using machine vision system. *Trans. IChemE(Part C)* 79: 223-231.
- [7] Dillon, W.R. and Goldstein, M. 1984. *Multivariate analysis methods and applications*. John Wiley and Sons, NY.
- [8] Gombar, V.K. and Enslein, K. 1991. A structure biodegradability relationship model by discriminant analysis. In *Chemical and Environmental Studies*, (J. Devillers and W. Karcher, eds.) pp. 377-414, Academic Publishers, The Netherlands.