Virtual Chemical Sensor For Classification Of Herb – *Orthosiphon stamineus* According To Its Geographical Origin

O.S. Chew[1], M.N. Ahmad[1], Z. Ismail[2], A.R. Othman[3]

[1]*School of Chemical Science*

[2]*School of Pharmaceutical Science*

[3]*Centre for Distance Education*
University Science Malaysia
11800 Minden Penang, Malaysia

Abstract - Herbal medicine is an important part of health care to a majority of the world's population. Herbal standardization is important to maintain products consistency and repeatability in their composition. Volatile components in herb - *Orthosiphon stamineus* were analyzed using a new chemical sensor concept based upon a fast Gas Chromatography (GC) with Surface Acoustic Wave (SAW) detector. The sensing principle is based on the injection of a complex sample headspace into the fast gas chromatography, creating a chromatogram of the unresolved gaseous mixture. After selecting the particular components, the resulting parameters – retention time and concentration of samples are treated with Pattern Recognition analysis namely Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Cluster Analysis (CA). This study was successfully classified *O. stamineus* according to their geographical origin. The developed method is very useful for reliable QA/QC for raw herbs and their products.

Keywords: Virtual Chemical Sensor, Electronic Nose; *Orthosiphon stamineus;* Fast Gas Chromatography; Chemometric; Pattern Recognition.

1. Introduction

Scientific research on Misai Kucing especially on species *Orthosiphon stamineus* had begun since 1970. According to Van der Venn [1], *O. stamineus* is applied as a medicinal plant because of the diuretic and bacteriostatic properties due to the content of potassium, inositol and lipophilic flavones in its leaves. Based on the preliminary report [2], further investigated the composition of the essential oil derived from commercial fresh leaf and stem namely Orthosiphon folium DAB 8 using Gas Chromatography-Mass Spectrometry (GC-MS). The result shown that β-caryophyllene, β-elemene, humulenem, β-bourbonene, 1-octen-3-ol and caryophyllene oxide were the main volatile organic compound.

0-7803-8102-5/03/$17.00 © 2003 IEEE

In Malaysia, research institute such as Institute of Medical Research and Kuala Lumpur Hospital of the Malaysian Health Ministry are participatory actively in conducting clinical study to prove the efficacy of *O. stamineus* in treating kidney stones disease. On the other hand, School of Pharmaceutical Science, University Science Malaysia contribute to the scientific research and development of *O. stamineus* by conducting the extraction, quality control, standardization, pharmacological and formulation research of *O. stamineus*.

Currently, the new version of electronic nose, which integrated the separation techniques of Gas Chromatography (GC) and a mass sensitive detector namely Surface Acoustic Wave (SAW) detector are being used [3]. In this study, the use of electronic nose technology is expanded to the field of natural and herbal product research. The developed method has the potential to be serve as an alternative analytical techniques for herbal analysis that is less time consuming, cost effective and easy to operate compared to conventional analytical techniques such as High Performance Liquid Chromatography (HPLC), Thin Layer Chromatography (TLC) and Gas Chromatography-Mass Spectrometry (GC-MS). stand-alone analytical technique for quantitative and qualitative analysis of chemical constituent in the raw and final herbal product.

Chemometrics is a term coined in 1972, which can be defined as the chemical discipline that uses mathematical, statistical and other methods employing formal logic to (a) design or select optimal measurement procedures and experiments and (b) to provide maximum relevant chemical information by analyzing chemical data [4]. A widely applied discipline of chemometrics is pattern recognition, which involves the classification and identification of samples. Its purpose is to develop a semiquantitative model that can be applied to the identification of unknown sample patterns [5]. As a conclusion, chemometrics analysis is used to analyze and interpret a cluster of raw data into knowledgeable information using statistic and mathematics model.

In this research, dried leaves of *O. stamineus* cultivated commercially in different geographical origin have been classified using a virtual chemical sensor based on fast Gas Chromatography (GC) with Surface Acoustic Wave (SAW) detector namely zNose™. The resultant instrumental data was processed with both unsupervised pattern recognition techniques that is Principle Component Analysis (PCA), Cluster Analysis (CA) and a supervised pattern recognition techniques namely Linear Discriminant Analysis (LDA) in order to classify the raw sample according to its place of origin.

2. Experimental

2.1 Materials

Samples *O. stamineus* (dried leaves) from five different geographical origins were collected from the respective distributor and the samples were named using alphabetical codes as shown in Table 1.

TABLE 1
The list of *O. stamineus* samples according to its geographical origin.

Code*	Distributor Name	Location	State	Country
SRKBPM	Shukor Rahman	Kepala Batas	Pulau Pinang	Malaysia
STJGCM	Shukri Talib	Jengka	Pahang	Malaysia
ZBPRAM	Zaibidi	Parit	Perak	Malaysia
NNPPDM	Nik Norma	Pasir Puteh	Klantn	Malaysia
NHPJI	Nusantara Herbs	Pulau Jawa	Jakarta	Indonesia

* Example SRKBPM: SR: distributor; KB: location; P: state; M: Country.

2.2 Sample preparation

An amount of dried samples were milled until fine powder. 0.1000g samples were placed in a 2ml headspace vial, which was then closed with a PTFE (Kimble Glass Inc.) septum cap. Fast GC/SAW electronic nose system namely zNose™, Model 7100 (Electronic Sensor Technology, California) was used to analyze volatile organic compound (VOC) from the headspace samples.

2.3 System component of GC/SAW electronic system

The system consists of a six-port, two-position valve; the loop trap; a sampling pump for pulling vapors into the loop trap; a source of clean helium for use as a carrier gas; the GC column, which is a short section of glass or metal capillary tubing ~0.25 mm in diameter; and the temperature controlled SAW vapor sensor [6].

2.4 Analytical conditions

Sensor measurements: Thermostating 30 minutes at 60°C with heating block (Cole Palmer Inc.). GC/SAW electronic nose operation condition were as follows; injection time 5 second, inlet temperature 180°C, valve temperature 120°C, detector temperature 40°C, ramp 40°C-100°C @ 10°C/sec, capillary column: DB-624, 1m x 0.25mm x 1μm, carrier gas: helium 10 ml/sec, data acquisition time 10 second. Triplicate measurements per vial were carried out for each different geographical origin samples.

2.5 Data transformation and data analysis

Data transformation was performed using MS Excel. First, a set of particular GC peak was chosen as "virtual sensor array" based on the corresponded GC profile. Then, the frequency data of each peak ("sensor") was calculated as the mean average frequency obtained from triplicate measurement. After that, the frequency data was standardized to zero mean and unit standard deviation by using SPSS 9.0 program. Finally, PCA, CA and LDA were also carried out using SPSS 9.0 software in order to classify the sample according to its geographical origin.

3. Results and discussion

3.1 Optimum virtual chemical sensor array selection

The prefect combination of GC direct-heated column and the SAW detector makes a virtual physical sensor. Although the system contains a single physical sensor, the compatible system software, namely MicroSense 3.6 can create hundreds of virtual chemical sensor based upon retention time windows. This means that each peak of the GC profile can be identified as a response of a single chemical sensor and at the same time correspond to only one analyte or chemical compound found in the sample. This approach was also reported by Dittmann [7] which assumes that each fragment ion obtains from mass spectrum characterized certain chemical compound. Typical profile chromatograms of five samples with numerical label represents selected virtual sensor response are shown in Figure 1 below.

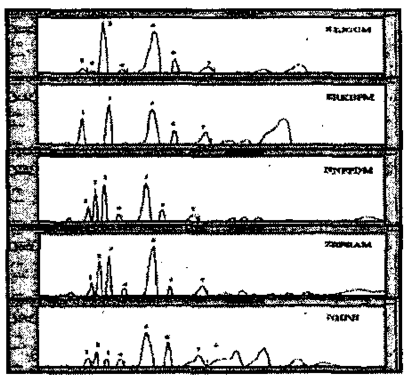

Figure 1 Profile chromatogram represent-ting an array of virtual chemical sensor, which characterized the different geographical origin samples respectively.

3.2 Classification by pattern recognition

3.2.1 Image olfactory (VaporPrint™ Image)

3.2.2

According to the inventor of the GC/SAW electronic nose system [8], image olfactory is a high-resolution (500 pixel) two-dimensional visually recognizable images, which can also quantify the strength of each chemical within a fragrance. The image is a closed polar plot of the odor amplitude (SAW detector frequency) with radial angles representing sensor. A brief conclusion can be drawn by making comparison among the vapor image shown in Figure 2. Hence, *O. stamineus* from different origin represents its own aroma pattern. So the unknown samples can be easily classified according to its origin by making comparison with the vapor image of reference sample. But this approach is not reliable when the vapor image look similar to each other as shown by ZBPRAM and NNPPDM samples. In this situation, high probability of misclassification can happen.

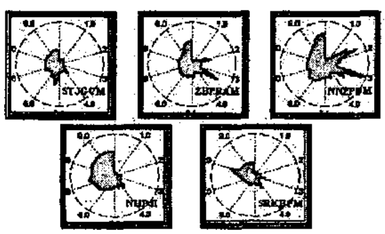

Figure 2 Image olfactory of different geographical origin *O. stamineus* samples.

Due to the above state limitations, chemometric approaches such as unsupervised pattern recognition techniques namely principle component analysis (PCA) and cluster analysis (CA) were investigated. Besides that, linear discriminant analysis (LDA) that is a supervised pattern recognition technique is also being applied in this study.

3.3 Chemometrics classification

3.3.1 Data pretreatment

According to Massart et al. [4], if the distribution of variables as in this study the frequency data are not normal but severely skewed, then the reliable or successful results cannot be obtained from most multivariate statistical analyses. On the other hand, Miller et al. [9] stressed that decision must be made whether raw data or standardized data (min=0 and standard deviation=1) is used for data analysis before PCA and LDA is carried out. Based on the above statement, data frequency from the instrument analysis is standardized to obtain equal weight of all the variables. If no data pretreatment is done, the zero reading in sensor 2 and sensor 4 shown by SRKBPM samples account for such a large variance and false classification may occur.

3.3.2 Principle component analysis (PCA)

Figure 3 shows the scatter plot of the standardized frequency data in two dimensions. The first two principle component represent 67.39% of the total variance (PC1 = 41.58% and PC2 = 25.80%). The two principle components are independent. A straight line passing through the data points represents a linear combination of the corresponding variables.

A good separation between NHPJI samples from the others samples is obtained. This observation explain the fact that the cultivated area of herbal medicines is a controlling factor in the quality of the herb due to the different growing conditions. Classification of NNPPDM and ZBPRAM samples indicated the volatile composition of the both samples are similar and do not differ enough to make a good separation [10].

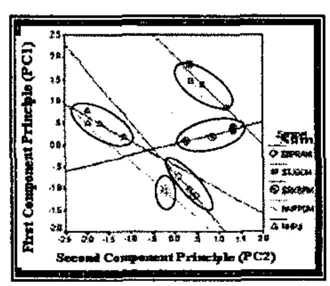

Figure 3 Principle component analysis of the virtual sensor array responses to *O. stamineus* samples of different geographical origin.

As a result, PCA is not so effective for classification of the *O. stamineus* samples according to its geographical origin. Although Togari et al. [11] reported that PCA showed effectiveness for classification of tea samples according to its categories (fermented and unfermented) based on the GC profile. Thus, the classification power of LDA a supervised pattern recognition techniques is investigated.

3.3.3 Linear discriminant analysis (LDA)

This supervised pattern recognition techniques had been applied widely for the classification purposes. Martin et al. [12], have proven that the LDA method shows good classification and prediction capabilities of vegetable oils.

LDA when applied to classify *O. stamineus* samples based on its origin seem to give good classification results as shown in Figure 4. By using LDA, SRKBPM and STJGCM samples separated well on the negative side of the x-axis and the y-axis, which was not clearly classified by PCA. As a conclusion, this study finds that LDA is a tool more powerful compared to PCA in terms of classification. This is mainly because LDA selects direction, which gives maximum separation between the studied classes [4].

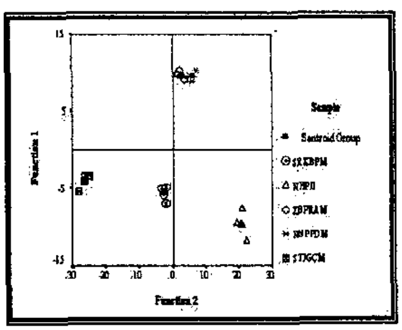

Figure 4 Discriminant plot of *O. stamineus* sample by LDA method.

3.3.4 Cluster analysis (CA)

CA was carried out using raw data obtained from the analytical instrument in order to study the capabilities of the selected virtual sensor array for classification of the samples based on geographical origin. This approach is able to, assign a group of objects to its respective classes so that similar object are in the same classes. The resultant dendogram gives extra information regarding the raw data obtained from instrumental analysis.

This study employs Average Linkage method and Euclidean distance as distance measure between objects. Euclidean distance was used because the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers. The dendogram (Figure 5) horizontal scale (0-25) give pictures of similarity and dissimilarity among samples. Samples from the same origin form individual clusters except SRKBPM sample as indicated by no. 19 form a cluster with ZBPRAM sample. The probable reason for this misclassification was mainly due to the zero reading sensor respond shown by SRKBPM samples.

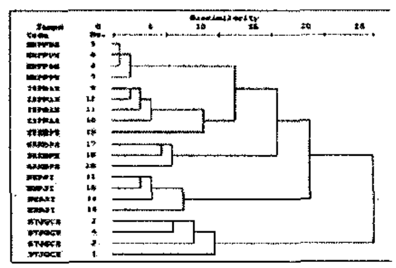

Figure 5 Dendrogram of cluster analysis on different geographical *O. stamineus* samples.

4. Summary

With the special integrated feature in the GC/SAW electronic nose system, for the first time E-nose can serve as an alternative analytical technique for herbal analysis that is less time consuming, cost effective and easy to operate compare to conventional analytical techniques such as High Performance Liquid Chromatography (HPLC), Tin Layer Chromatography (TLC) and Gas Chromatography-Mass Spectrometry (GC-MS). Chemometric pattern recognition applied to the selected optimum virtual sensor data from the GC profile is effective in classifying of *O. stamineus* samples according to its geographical origin. The combination of the chemometrics approach and GC/SAW electronic nose shows to be a promising analytical technique for herbal analysis. Further study is needed with some modifications in the analytical procedure that emphasized on quantification of the chemical constituent in *O. stamineus.*

5. Acknowlwdgement

The authors gratefully acknowledge the financial support from the Ministry of Science, Technology and Environment, Malaysia through IRPA No. 610629 for this research.

6. References

[1] X. Van deer Veen, Th.M. Makubard, J.H., Z. wazing, *Pharmaceutisch Weekblad,* vol. 114, pp. 965, 1979.

[2] S. Schmidt, R. Bos, "Volatile Of *Orthosiphon stamieus* Benth", *Progress in Essential Oil Research,* Walter de Gruyter & Co., Berlin: New York, pp. 93-97, 1986.

[3] M.N. Ahmad, E.M. Zaihidee, O.S. Chew, A.R. Othman, Z. Ismail, M.S. Hitam, A.Y. Md Shakaff, "Chemometric Classification of Herb – *Orthosiphon Stamineus* (Cat Whisker's)
According to Geographical Origin Using a New Chemical Sensor Based on a Fast GC", The 1st *International Meeting On Microsensors & Microsystems,* National Cheng Kung University, Tainan, Taiwan, 12-14 January 2003.

[4] D.L Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: A Textbook,* vol. 2, Elsevier: Amsterdam. 1998.

[5] H.Y. Aboul-Enein, R. Stefcin, G. Baiulesscu, *Quality & Reliability in Analytical Chemistry,* CRC Press UC, pp. 49-53, 2000.

[6] E.J. Staples, T. Matsura, , S. Viswanathan, "Real Time Environmental Screening Of Air, Water And Soil Matrices Using A Novel Field Portable GC/SAW System", *Environmental Strategies for the 21st Century, Asia Pacific Conference,* Singapore, 8-10 April 1998.

[7] B. Dittmann, S. Nitz, "Strategies For The Development Of Reliable QA/QC Method When Working With Mass Spectrometry-Based Chemosensory Systems", *Sensor and Actuators B.,* vol. 69, pp. 253-257, 2000.

[8] E.J. Staples, "Electronic Nose Simulation of Olfactory Response Containing 500 Orthogonal Sensors in 10 Seconds", *Proceeding of the 1999 IEEE Ultrasonics Frequency Control and Ferroelectrics Symposium,* Lake Tahoe, CA, Oct 18-21 1999.

[9] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry,* 4th ed. Prentice Hall, 2000.

[10] I.J Kosir, M. Kocjancic, N. Ogrinc, J. Kidric, "Use Of SNIF-NMR And IRMS In Combination With Chemometric Method For The Determination Of Chaptalisation And Geograhical Origin Of Wines (The Example Of Slovenian Wines)", *Analytical Chimica Acta,* Vol 429, pp. 195-206, 2001.

[11] N. Togari, A. Kobayashi, T. Aishima, "Pattern Recognition Applied To Gas Chromatographic Profiles Of Volatile Components In Three Tea Categories", *Food Research International.* vol. 28(5), pp. 495-502, 1995.

[12] Y.G. Martin, M.C. Cerratp Oliveros, J.L. Perez Pavon, C.F. Pinto, B.M. Cordero, "Electronic Nose Based On Metal Oxide Semiconductor Sensor And Pattern Recognition Techniques : Characterisation Of Vegetable Oils", *Analytical Chimica Acta.* vol. 449, pp. 69-80, 2000.