

Article

Odor Discrimination by Similarity Measures of Abstract Odor Factor Maps from Electronic Noses

Weiqing Guo ¹, Haohui Kong ², Junzhang Wu ² and Feng Gan ^{1,*} ¹ School of Chemistry, Sun Yat-Sen University, Guangzhou 510275, China; guowqing@mail2.sysu.edu.cn² Technology Center, China Tobacco Guangdong Industrial Co., Ltd., Guangzhou 510385, China; konghh@gdzygy.com (H.K.); wujunzhang@gdzygy.com (J.W.)

* Correspondence: cesgf@mail.sysu.edu.cn

Received: 23 July 2018; Accepted: 7 August 2018; Published: 13 August 2018



Abstract: The aim of this study is to improve the discrimination performance of electronic noses by introducing a new method for measuring the similarity of the signals obtained from the electronic nose. We constructed abstract odor factor maps (AOFMs) as the characteristic maps of odor samples by decomposition of three-way signal data array of an electronic nose. A similarity measure for two-way data was introduced to evaluate the similarities and differences of AOFMs from different samples. The method was assessed by three types of pipe and powder tobacco samples. Comparisons were made with other techniques based on PCA, SIMCA, PARAFAC and PARAFAC2. The results showed that our method had significant advantages in discriminating odor samples with similar flavors or with high VOCs release.

Keywords: electronic nose; abstract odor factor maps; similarity measure; odor discrimination

1. Introduction

In Nature, mammals discriminate odor through a complex process. First, the olfactory sensory neurons detected odor molecules, then axons on the neurons transmit the signal to olfactory bulb, and finally the olfactory bulbs process the signal to obtain the information of the odor [1]. Inspired by the odor discrimination process of mammals, researchers have developed artificial olfactory systems called electronic nose systems that contains perception, signal processing and recognition sections [2,3]. Although the discrimination ability of electronic nose systems is far beneath that of the mammalian olfactory system, these systems have been quickly applied to odor discrimination in many fields, such as agriculture [4–6], fishery [7–9], the food industry [10–12], disease diagnosis [13–15], environmental monitoring [16–18] and chemical safety [19,20] because of their objectivity, stability and durability. Compared with the traditional odor analysis techniques, such as gas chromatography and its hyphenated techniques [21,22], electronic nose systems show good application prospects because they are fast and sensitive, need simple pre-processing and operation steps, and can obtain overall information of the volatile components in samples.

However, commercial electronic nose systems usually have low selectivity [23] and cross-sensitivity [24], which limits their discrimination performance. Many data processing methods have been introduced to improve the discrimination performance of electronic noses [23,25–32]. The traditional data analysis methods for electronic noses are mainly based on bilinear models, such as PAC, SIMCA, DFA, KNN, SVM, etc. [25–30] Later, methods based on trilinear models were also applied to electronic nose data analysis, such as PARAFAC, PARAFAC2 and MOLMAP [23,31,32]. However, their performances were not quite satisfactory when electronic nose was applied to odor samples with high similarity or complexity. The reason could be that there were differences between the real response models of the sensor arrays and the mathematical models of these methods. Recently,

researchers have attempted to solve the problem through new techniques, such as feature extraction (selection) [33–36], nonlinear modification [35,37], interference suppression algorithms [24] and so on. These techniques, to some extent, can improve the discrimination performance of electronic noses, but do not essentially solve the problem. New data processing methods are needed.

In this paper, we establish a new data processing method for the discrimination of similar or complex odor samples. A novel signal model based on the true response mechanism of a metal oxide semiconductor (MOS) sensor array had been developed in our previous work and had shown success in the discrimination of perfume samples [38]. Here, we develop a new algorithm for the decomposition of this signal mode, and further propose a conception of abstract odor factor maps (AOFMs). An AOFM is a two-way matrix that is reconstructed by the abstract factors decomposed from a three-way electronic nose signal data array. Thus, it can characterize the main odorous substances in the sample and be used as a characteristic map of an odor. A similarity measure method [39] for two-way data is introduced to evaluate the similarities and differences between the AOFMs of samples. The new method is applied to three types of pipe and powder tobacco samples, the results were compared with those of PCA, SIMCA, PARAFAC and PARAFAC2. The results showed that the new method outperformed the aforementioned methods, showing significant advantages in discriminating complex or similar odor samples.

2. Methodology

2.1. Algorithm for the Decomposition of Signal Model

We assume that an electronic nose system is made up of k sensors. When this system is used to measure n samples and the measuring time is t for each sample, a three-way data array $\underline{\mathbf{R}}_{t \times k \times n}$ is obtained $\underline{\mathbf{R}}_{t \times k \times n}$. According to our previous study [38], each slice of $\underline{\mathbf{R}}_{t \times k \times n}$ can be expressed as a two-way data $\mathbf{R}_{t \times k, i}$, and be decomposed as follows:

$$\mathbf{R}_{t \times k, i} = \mathbf{C}_{t \times k} \mathbf{N}_{p \times (ck), i} \mathbf{\Gamma}_{(ck) \times k} \quad (1)$$

Here, c is the number of abstract odor molecules and p is the number of factors. $\mathbf{C}_{t \times p}$ is the concentration profiles of absorbed mass; $\mathbf{N}_{p \times (ck), i}$ is a scaling matrix indicating the absorbed-amount ratios of sensors. $\mathbf{\Gamma}_{(ck) \times k}$ is the matrix of odor characteristics which is band-diagonal matrix.

In our previous work, we temporarily decompose by a simple alternating least squares. Here we propose a new decomposition algorithm inspired by parallel factor analysis (PARAFAC).

(1) Step 1: obtain an initial $\mathbf{C}_{t \times p}$. The slabs of $\underline{\mathbf{R}}_{t \times k \times n}$ are stringed out horizontally as follows:

$$\mathbf{R}_{\text{col}} = [\mathbf{R}_{t \times k, 1} \mathbf{R}_{t \times k, 2} \dots \mathbf{R}_{t \times k, n}] \quad (2)$$

Through principal component analysis of \mathbf{R}_{col} , the top p score vectors $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p$ are used to construct $\mathbf{C}_{t \times p}$:

$$\mathbf{C}_{t \times p} = [\mathbf{t}_1 \mathbf{t}_2 \dots \mathbf{t}_p] \quad (3)$$

(2) Step 2: obtain an initial $\mathbf{\Gamma}_{(ck) \times k}$. The slabs of $\underline{\mathbf{R}}_{t \times k \times n}$ are stringed out vertically as follows:

$$\mathbf{R}_{\text{row}} = \begin{bmatrix} \mathbf{R}_{t \times k, 1} \\ \mathbf{R}_{t \times k, 2} \\ \vdots \\ \mathbf{R}_{t \times k, n} \end{bmatrix} \quad (4)$$

The first c loadings $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_c$ that are obtained from principal component analysis of \mathbf{R}_{row} are used to construct $\mathbf{\Gamma}_{(ck) \times k}$ as follows:

$$\mathbf{\Gamma}_{(ck) \times k} = \begin{bmatrix} \begin{pmatrix} p_1^{(1)} \\ p_1^{(2)} \\ \vdots \\ p_1^{(c)} \end{pmatrix} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \begin{pmatrix} p_2^{(1)} \\ p_2^{(2)} \\ \vdots \\ p_2^{(c)} \end{pmatrix} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \begin{pmatrix} p_k^{(1)} \\ p_k^{(2)} \\ \vdots \\ p_k^{(c)} \end{pmatrix} \end{bmatrix} \quad (5)$$

(3) Step 3: calculate $\mathbf{N}_{p \times (ck),i}$ by $\mathbf{C}_{t \times p}$ and $\mathbf{\Gamma}_{(ck) \times k}$:

$$\mathbf{N}_{p \times (ck),i} = \left((\mathbf{C}_{t \times p})^T (\mathbf{C}_{t \times p}) \right)^{-1} (\mathbf{C}_{t \times p})^T \mathbf{R}_{t \times k,i} \left(\mathbf{\Gamma}_{(ck) \times k} \right)^T \left(\left(\mathbf{\Gamma}_{(ck) \times k} \right) \left(\mathbf{\Gamma}_{(ck) \times k} \right)^T \right)^{-1} \quad (6)$$

(4) Step 4: calculate $\mathbf{C}_{t \times p}$ by $\mathbf{N}_{p \times (ck),i}$ and $\mathbf{\Gamma}_{(ck) \times k}$:

$$\mathbf{N}\mathbf{\Gamma} = [\mathbf{N}_{p \times (ck),1} \mathbf{\Gamma}_{(ck) \times k} \mathbf{N}_{p \times (ck),2} \mathbf{\Gamma}_{(ck) \times k} \dots \mathbf{N}_{p \times (ck),n} \mathbf{\Gamma}_{(ck) \times k}] \quad (7)$$

$$\mathbf{C}_{t \times p} = \mathbf{R}_{\text{col}}(\mathbf{N}\mathbf{\Gamma})^T \left(\mathbf{N}\mathbf{\Gamma}(\mathbf{N}\mathbf{\Gamma})^T \right)^{-1} \quad (8)$$

(5) Step 5: calculate $\mathbf{\Gamma}_{(ck) \times k}$ by $\mathbf{C}_{t \times p}$ and $\mathbf{N}_{p \times (ck),i}$:

$$\mathbf{C}\mathbf{N} = \begin{bmatrix} \mathbf{C}_{t \times p} \mathbf{N}_{p \times (ck),1} \\ \mathbf{C}_{t \times p} \mathbf{N}_{p \times (ck),2} \\ \vdots \\ \mathbf{C}_{t \times p} \mathbf{N}_{p \times (ck),n} \end{bmatrix} \quad (9)$$

$$\mathbf{\Gamma}_{(ck) \times k} = \left((\mathbf{C}\mathbf{N})^T \mathbf{C}\mathbf{N} \right)^{-1} (\mathbf{C}\mathbf{N})^T \mathbf{R}_{\text{row}} \quad (10)$$

(6) Step 6: calculate residual sum of squares (SSR):

$$\text{SSR} = \sum_{i=1}^n \left(\mathbf{R}_{t \times k,i} - \mathbf{C}_{t \times p} \mathbf{N}_{p \times (ck),i} \mathbf{\Gamma}_{(ck) \times k} \right)^2 \quad (11)$$

(7) Repeat Steps (3), (4), (5) and (6) until SSR reaches preset value.

In the decomposition process, three constraints are implemented (1) nonnegative constraints are implemented to matrices \mathbf{C} , \mathbf{N} and $\mathbf{\Gamma}$; (2) normalization in unit length and unimodal constraint are implemented to each column of matrix \mathbf{C} ; (3) a band-diagonal constraint is implemented to matrix $\mathbf{\Gamma}$.

2.2. The Construction of AOFMs

When the response matrix $\mathbf{R}_{t \times p, i}$ of an odor sample is decomposed into p factors according to the Equation (1), its p odor maps can be re-constructed in following way:

$$\mathbf{R}_j^* = \mathbf{c}_j \mathbf{n}_j^T \Gamma_{(ck) \times k}, j = 1, 2, \dots, p \quad (12)$$

where, \mathbf{R}_j^* is the j th odor map, \mathbf{c}_j is the j th column of matrix $\mathbf{C}_{t \times p}$, \mathbf{n}_j^T is the transpose of the j th row of matrix $\mathbf{N}_{p \times (ck), i}$. A two-way matrix \mathbf{R}^* , the AOFMs, is then constructed as follows:

$$\text{AOFM} \sim \mathbf{R}^* = \begin{bmatrix} \mathbf{R}_1^* \\ \mathbf{R}_2^* \\ \vdots \\ \mathbf{R}_p^* \end{bmatrix} \quad (13)$$

2.3. Similarity Measure of AOFMs

One of the authors of this paper had put forward a similarity measure method for two-way data [39]. A brief explanation of the method is given here. If the AOFMs of two odor samples have the forms of \mathbf{R}_A^* and \mathbf{R}_B^* , their differences can be calculated as follows:

$$\Delta \mathbf{R}^* = \mathbf{R}_A^* - \mathbf{R}_B^* \quad (14)$$

A statistic is constructed as follows, which is the mean value of $\Delta \mathbf{R}^*$:

$$T = \frac{1}{m \times n} \sum_{j=1}^n \sum_{i=1}^m \Delta \mathbf{R}^*(i, j) \quad (15)$$

where, m and n are the row and column of $\Delta \mathbf{R}^*$. When T is known, a hypothesis test and its posterior probability are expressed as follows:

$$\begin{cases} H_0 & \Delta \mathbf{R}^* = \mathbf{R}_A^* - \mathbf{R}_B^* = 0 \\ H_1 & \Delta \mathbf{R}^* = \mathbf{R}_A^* - \mathbf{R}_B^* \neq 0 \end{cases} \quad (16)$$

$$POR = \frac{p(H_0|T)}{p(H_1|T)} = \left[\frac{p(H_0)}{p(H_1)} \right] \left[\frac{p(T|H_0)}{p(T|H_1)} \right] \quad (17)$$

Here, POR is the posterior odds ratio. Equation (17) can be converted to $POR = \alpha LR$. The mathematical expressions of α and LR are Equations (S1) and (S2) (supporting information), respectively. Parameter LR is the likelihood ratio of two samples and α represents prior probabilities. LR can be calculated by Equation (S4) and α can be calculated by the AOFMs of samples of a training set, which is described in reference [39] or supporting information. To discriminate an odor sample in prediction set, its AOFMs should be compared with the mean AOFMs in the training set. If $POR < 1$, the two odor samples are statistically different; If $POR > 1$, they are considered with same odor. In this way, odor samples can be discriminated.

For the prediction of an unknown samples, we should firstly calculate the T between the AOFMs of unknown samples and the mean AOFMs of training set samples according to Equations (14) and (15). When T is known, the parameter LR can be calculated by Equation (S4). Since the parameter α of the training set samples is known, the values of POR between the unknown samples and training set samples are calculated according to $POR = \alpha LR$. Finally, the similarities between the samples are judged by the value of POR .

3. Materials and Methods

3.1. Instruments

All the data were obtained from an α Fox-4000 electronic nose equipped with Alpha Multi Organoleptic System (Alpha MOS, Toulouse, France). The system contains 18 MOS sensors that are placed in three chambers, a HS-100 auto-sampler, and an AG2301 high pure air generator.

3.2. Sample and Measurement Condition

Three types of pipe tobacco and tobacco powder samples with similar flavor were provided by the Technology Center of Tobacco Guangdong Industrial Co., Ltd. (Guangzhou, China). The pipe tobacco samples are labeled as Pipe Tobacco I, Pipe Tobacco II and Pipe Tobacco III, while powder samples are labeled Tobacco Smalls I, Tobacco Smalls II and Tobacco Smalls III. Each group of the samples included 18 samples, and hence in total there were 54 Pipe Tobacco and 54 Tobacco Smalls.

The measurement conditions were optimized by experiments. A certain amount of pipe/powder was taken into a 10 mL glass vial and then sealed as a sample. The sample was placed in HS-100 auto-sampler and equilibrated for 600 s. The equilibrating temperature for Pipe Tobacco was 60 °C, and for Tobacco Smalls was 35 °C. After that, 2500 μ L of the headspace gas in the vial was injected into the syringe by a sampling pump at the speed of 350 mL/min, and then injected into the electronic nose system along with high pure air which acted as carrier gas at the speed of 500 mL/min. The response of system in first 100 s was used for data analysis. It was recorded every 1 s. The sensors were cleaned up by high pure air for 1080 s after each measurement.

3.3. Data Processing

Pre-processing of data was made before the signal model was applied to the data. The pre-processing methods are shown in Equations (S5) and (S6) (Supporting Information). After the measurement of each sample, a two-way data with the size of 100×18 (time \times sensor) was obtained. Thus, 18 samples provided 18 data arrays for each group of samples. Finally, the 18 two-way data arrays were arranged into a three-way data with the size of $100 \times 18 \times 18$ (time \times sensor \times sample).

To obtain the optimal result of the model, the parameters c and p should be optimized. The optimized range of parameter c was from 1 to 3, and p was from 1 to 5. The data were decomposed based on the signal model by setting different c and p . The sum of squares of residual (SSR) was the optimization criterion. The data analysis for PCA and SIMCA were performed using the AlphaSoft, Version 12.42 (Alpha MOS). The data analysis for other methods were implemented using Octave, Vision 4.0.3.

4. Results and Discussion

4.1. The AOFMs of Tobacco Smalls and Pipe Tobacco

The pre-processing method and proposed decomposition algorithm had been applied to the signal data of Pipe Tobacco and Tobacco Smalls samples. The decomposition results with different parameters c and p are shown in Tables S1 and S2. It showed that the optimum value of c for all tobacco samples was 2. Meanwhile, the optimal p value of Pipe Tobacco samples was 2, while the optimal p value of Tobacco Smalls samples was 4. The data of Pipe Tobacco and Tobacco Smalls samples were decomposed with the optimal value of parameters, and we obtained their odor maps which were shown in Figure 1.

It should be explained that the number of odor maps depends on the value of model parameter p (see Equation (12)). As the p value of Pipe Tobacco samples was 2, each Pipe Tobacco sample had obtained two odor maps which are shown in Figure 1(A-1–A-3). For Tobacco Smalls samples, the value of p was 4 and thus all of them gave four odor maps which are shown in Figure 1(B-1–B-3). We noted that odor maps of the Tobacco Smalls samples were more than Pipe Tobacco. This result can

be explained. Tobacco Smalls were powder samples and thus they had larger specific surface area than Pipe Tobacco samples which were filamentous. Larger specific surface area was more conducive to the release of volatile organic compounds (VOCs). Therefore, the odors of Tobacco Smalls samples were more complex than Pipe Tobacco and need more odor maps to represent their feature information.

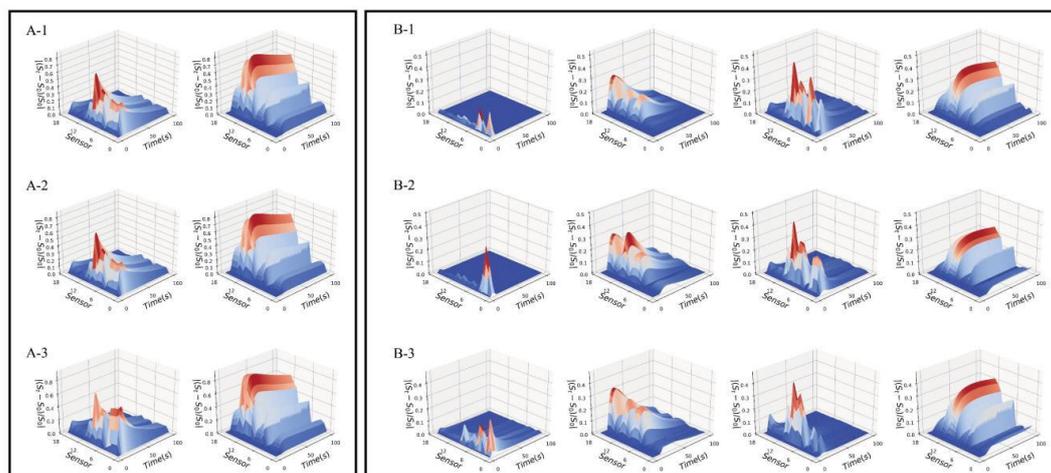


Figure 1. The odor maps of three types of Pipe Tobacco (A) and Tobacco Smalls (B). (A-1) Pipe Tobacco I; (A-2) Pipe Tobacco II; (A-3) Pipe Tobacco III; (B-1) Tobacco Smalls I; (B-2) Tobacco Smalls II; (B-3) Tobacco Smalls III.

We could distinguish the different odor samples by visually comparing their odor maps one by one. However, such a strategy was relatively time-consuming and lacked any objective criteria. Here, we reconstruct the odor maps into a two-way AOFM to comprehensively express the odor characteristics of each sample. The AOFMs of all Pipe Tobacco and Tobacco Smalls samples are shown in Figure 2, and we would use a similarity measure to evaluate their differences instead of visual observation.

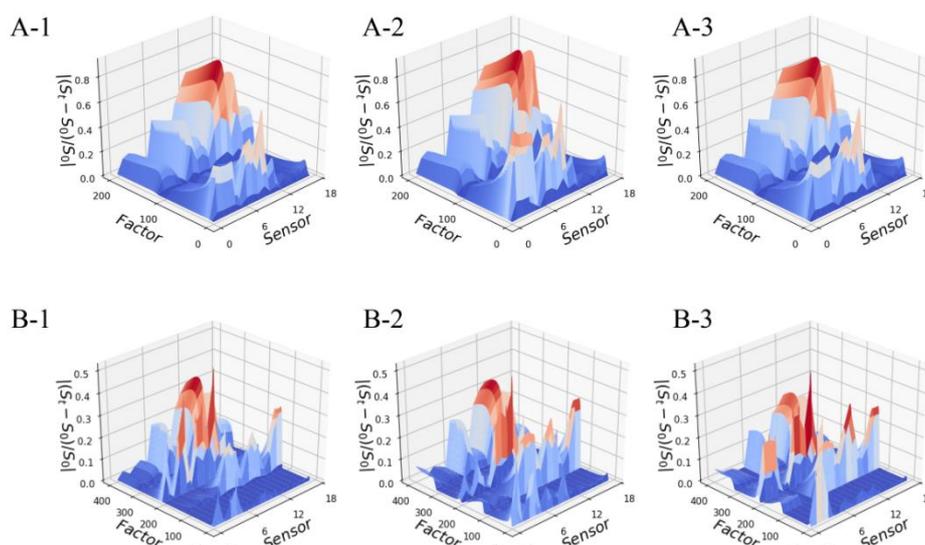


Figure 2. The two-way AOFM of three Pipe Tobacco samples (A) and three Tobacco Smalls samples (B): (A-1) Pipe Tobacco I; (A-2) Pipe Tobacco II; (A-3) Pipe Tobacco III; (B-1) Tobacco Smalls I; (B-2) Tobacco Smalls II; (B-3) Tobacco Smalls III.

4.2. Similarity Measure of AOFM

In order to evaluate the differences between the odors of the Pipe Tobacco samples, we set the AOFMs of each group of tobacco samples as the training set, and AOFMs of the other two groups as the prediction set. The similarities of AOFMs between prediction and training set samples were calculated (represented by the values of *POR*) as the criteria for the discrimination of their odor. The similarity measure results of AOFMs of Pipe Tobacco and Tobacco Smalls are summarized in Table 1.

Table 1. The similarity measure results of AOFMs of Pipe Tobacco and Tobacco Smalls.

No.	Training Set	Predicted Set	The Value of <i>POR</i>	Whether Be Separated
Pipe Tobacco	I	II	$2.87 \times 10^{-209} \sim 9.14 \times 10^{-88}$	Yes
		III	$\rightarrow 0.00^1$	Yes
	II	I	$5.32 \times 10^{-59} \sim 1.81 \times 10^{-14}$	Yes
		III	$\rightarrow 0.00^1$	Yes
	III	I	$\rightarrow 0.00^1$	Yes
		II	$\rightarrow 0.00^1$	Yes
Tobacco Smalls	I	II	$\rightarrow 0.00^1$	Yes
		III	$\rightarrow 0.00^1$	Yes
	II	I	$\rightarrow 0.00^1$	Yes
		III	$\rightarrow 0.00^1$	Yes
	III	I	$\rightarrow 0.00^1$	Yes
		II	$\rightarrow 0.00^1$	Yes

¹ $\rightarrow 0.00$ means the value of *POR* is infinitely close to 0.00.

When the AOFMs of Pipe Tobacco I were set as the training set, we got $POR < 1$ for both Pipe Tobacco II and III. This meant that Pipe Tobacco II and III were different from I, and thus Pipe Tobacco I samples can be distinguished from them. Pipe Tobacco II and III also can be discriminated by the same method. We noted when the AOFMs of Pipe Tobacco I were used as the training set, the *POR* values of AOFMs of Pipe Tobacco III were less than those of Pipe Tobacco II. Similarly, when the AOFMs of Pipe Tobacco II were used as the training set, the *POR* values of AOFMs of Pipe Tobacco III were also much less than those of Pipe Tobacco I. This finding supported the notion that the AOFMs of Pipe Tobacco I and II were similar to those of Pipe Tobacco III. Actually, Pipe Tobacco I, II and III were tobaccos with strong flavor. Pipe Tobacco I and II were from the same series while Pipe Tobacco III was from a different series. Hence, the flavors of Pipe Tobacco I and II were more similar than III and the similarity measure result was consistent with this fact.

The same method was also applied to calculate the *POR* values of the AOFMs of the Tobacco Smalls samples. We found that the *POR* values obtained from the six settings were also less than 1 and infinitely close to 0.00. This indicated that there were great differences between the prediction and the training samples. The results indicated that these three groups of Tobacco Smalls samples can be identified correctly from each other. Cellini et al. [40] noted that the discrimination performance of electronic nose might be weakened when it was applied to samples with high VOCs release. In this case, our method can effectively solve this problem.

4.3. Other Methods

In order to assess the advantages and disadvantages of our new method, the result is compared with those from PCA, SIMCA, PARAFAC and PARAFAC2.

4.3.1. PCA

PCA is a pattern recognition method contained in the α Fox4000 software. It had been used for analyzing measured data of tobacco samples, and the result is shown in Figure 3.

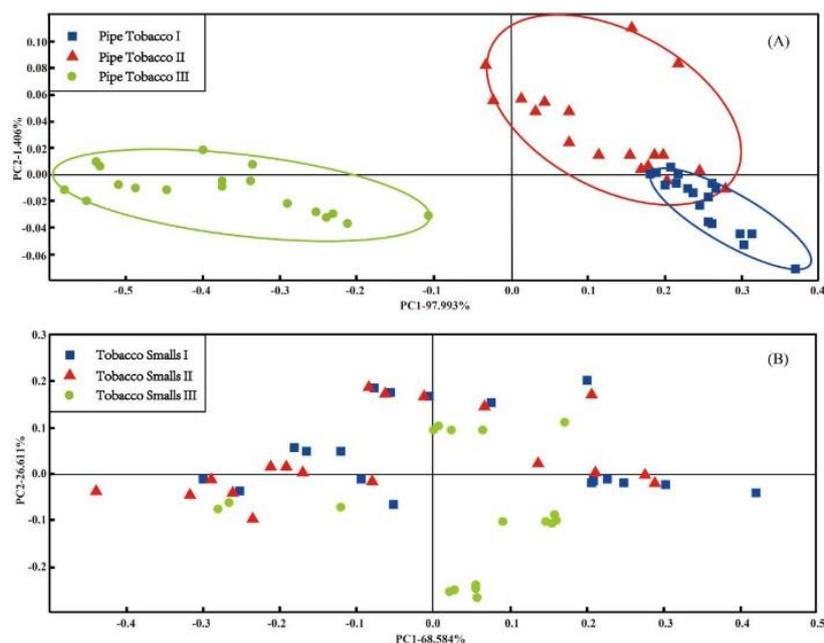


Figure 3. The PCA score plots of three types of Pipe Tobacco samples (A) and Tobacco Smalls (B).

Figure 3A is the PCA score plot of the Pipe Tobacco samples. It shows that there were certain differences among the odors of the three groups Pipe Tobacco samples. The points of Pipe Tobacco I and Pipe Tobacco II were obviously overlapped with each other, but Pipe Tobacco III points were separated from the other two Pipe Tobacco groups. This result implied that PCA can effectively discriminate Pipe Tobacco III, but cannot discriminate Pipe Tobacco I and II as they were more similar. For the Tobacco Smalls samples, the result is quite unsatisfactory. The three groups of Tobacco Smalls points were overlapped completely, and none of them can distinguished from each other (Figure 3B). The high VOCs release prevented an effective electronic nose discrimination in these results. According to Cellini et al. [40], too many VOCs released from Tobacco Smalls sample may be prone to producing a high background noise. This situation finally resulted that the samples of three groups Tobacco Smalls cannot be distinguished in the PCA score plot.

4.3.2. SIMCA

The SIMCA method is another data analysis method contained in α Fox4000 software which is used for cluster analysis. Figure 4(A-1–A-3) are the SIMCA results of Pipe Tobacco while Figure 4(B-1–B-3) are the results of the Tobacco Smalls. The area with blue background in Figure 4 is the identification area. When prediction samples fall in the identification area, they would be considered the same as the training samples, otherwise they would be considered as different odors. In Figure 4(A-1), the samples of Pipe Tobacco I were used as training set, and the samples of Pipe Tobacco II and III were used as the predicted set. We can find that some of Pipe Tobacco II samples fell in the identification area, and were wrongly recognized as Pipe Tobacco I. Meanwhile, all samples of Pipe Tobacco III fell away from the identification area, and hence were correctly discriminated from Pipe Tobacco I. The training samples of Figure 4(A-2) were Pipe Tobacco II while prediction samples were Pipe Tobacco I and III. The result showed that all Pipe Tobacco I samples were wrongly recognized as Pipe Tobacco II, but the Pipe Tobacco III samples were fell outside the identification area and were correctly distinguished from Pipe Tobacco II. When Pipe Tobacco III samples were used as training set (Figure 4(A-3)), both the Pipe Tobacco I and II samples stayed away from the identification area and were correctly distinguished. Therefore, we can find the SIMCA results were similar to the PCA ones, and inferior to the similarity measure of AOFMs. The Pipe Tobacco III samples can be distinguished from the other two groups of Pipe Tobacco, but Pipe Tobacco I and II samples cannot be distinguished from each other by the SIMCA

method due to their higher similarity. The training samples of Figure 4(B-1–B-3) were Tobacco Smalls I, II and III, respectively. From these three subfigures, we can see that almost all the samples had fallen into the recognition area and were considered as the same samples. Particularly in Figure 4(B-1,B-2), all points of samples were disordered. The points representing the same group of Tobacco Smalls were diffuse, and those representing different groups were mixed. In Figure 4(B-3), though all samples also were in the recognition area and misjudged as the same samples, samples of Tobacco Smalls III were still clustered and away from those of Tobacco Smalls I and II. Thus, even though SIMCA also cannot effectively discriminate between these three groups of Tobacco Smalls, its results were better than PCA. The reason may be that SIMCA is more effective in extracting feature information as it is a supervised learning method and PCA is an unsupervised one.

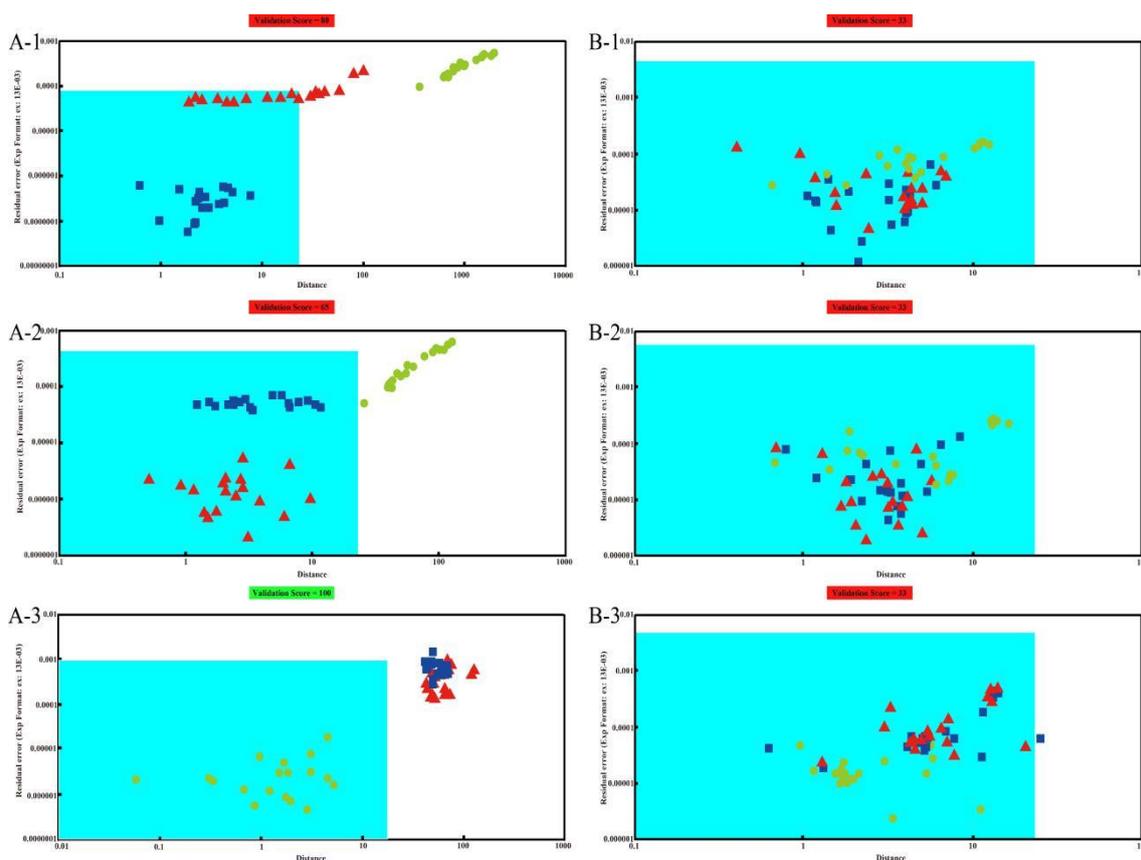


Figure 4. The SIMCA results of three types of Pipe Tobacco samples (A) and Tobacco Smalls (B). The training sets are Pipe Tobacco I (A-1), Pipe Tobacco II (A-2), Pipe Tobacco III (A-3), Tobacco Smalls I (B-1), Tobacco Smalls II (B-2), and Tobacco Smalls III (B-3), respectively.

4.3.3. PARAFAC and PARAFAC2

Considering that the data of electronic nose can be expressed as three-way data array, three-way resolution methods, such as parallel factor analysis (PARAFAC), are also applicable options for analyzing electronic nose data. In this paper, PARAFAC and PARAFAC2 were used to discriminate Pipe Tobacco and Tobacco Smalls samples, and their number of components were optimized. Figure S1 shows the trends of SSR versus the number of components when PARAFAC and PARAFAC2 were implemented for both the Pipe Tobacco and Tobacco Smalls samples. The results show that three principal components were appropriate for these PARAFAC and PARAFAC2 models. Figure 5(A-1,A-2) are the 3D score plots of PARAFAC and PARAFAC2 for Pipe Tobacco samples. The figures showed that Pipe Tobacco III samples were clustered but were separated from the Pipe Tobacco I and II ones. However, for the samples of Pipe Tobacco I and II, some of the points were overlapped.

These results indicated that both PARAFAC and PARAFAC2 could distinguish Pipe Tobacco III but cannot distinguish Pipe Tobacco I and II which were more similar. The results of PARAFAC and PARAFAC2 for Tobacco Smalls were even worse. Their 3D score plots are shown in Figure 5(B-1,B-2). The samples of the same groups of Tobacco Smalls were diffused and those of different groups were mixed in these two subfigures. Thus, PARAFAC and PARAFAC2 also had difficulties to distinguish Tobacco Smalls samples as they were more complex.

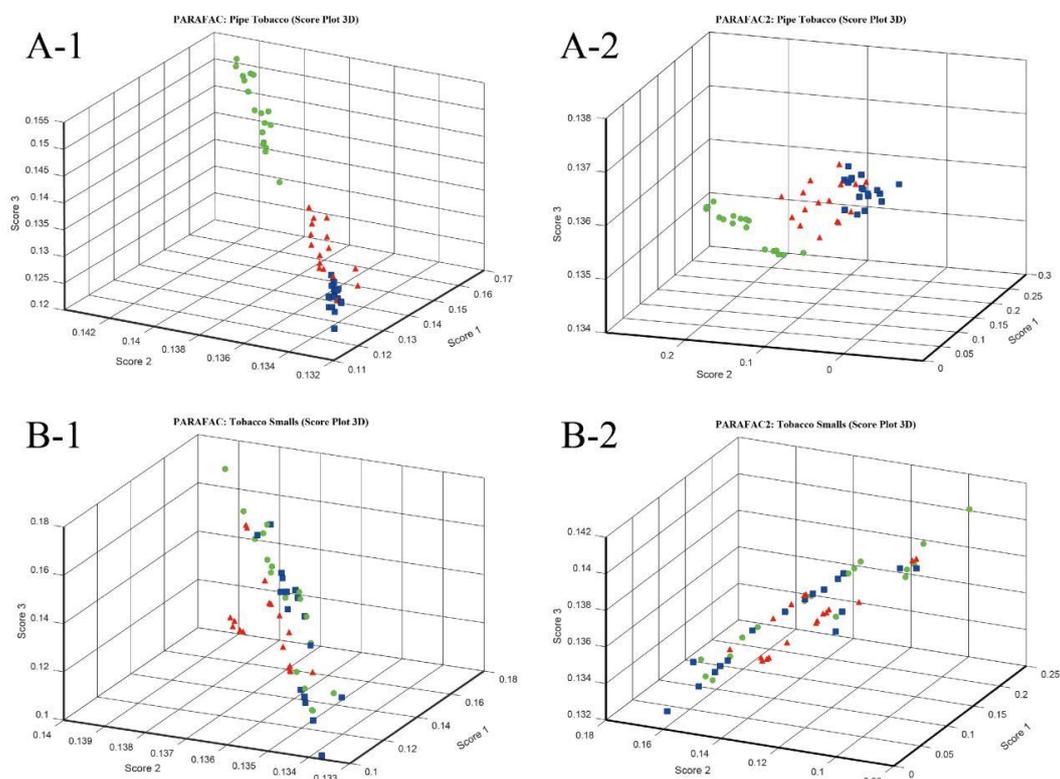


Figure 5. 3D score plot with top three components of Pipe Tobacco and Tobacco Smalls by PARAFAC and PARAFAC2. (A-1) Pipe Tobacco, PARAFAC; (A-2) Pipe Tobacco, PARAFAC2; (B-1) Tobacco Smalls, PARAFAC; (B-2) Tobacco Smalls, PARAFAC2.

4.4. Comparison

Table 2 lists the discrimination performance of the five methods. We can find that similarity measure of AOFMs outperformed PCA, SIMCA, PARAFAC and PARAFAC2. Comparing with the results of the aforementioned methods, the similarity measure of AOFMs can improve the discrimination performance of electronic noses in their applications to samples with similar flavors (Pipe Tobacco I and II) or with high VOCs releases (Tobacco Smalls samples).

Table 2. The discrimination performance of the five methods.

Label		Separation of Samples ¹				
		Similarity Measure of AOFM	PCA	SIMCA	PARAFAC	PARAFAC2
Pipe Tobacco	I-II	+	-(+)	-	-(+)	-(+)
	I-III	++	++	++	++	++
	II-III	++	++	++	++	++
Tobacco Smalls	I-II	++	-	-	-	-
	I-III	++	-	-	-	-
	II-III	++	-	-	-	-

¹ Separation of different tobacco samples evaluated from results of five methods. -: No separation; -(+): samples are partially separated; +: samples are separated; ++: samples are well separated.

5. Conclusions

An effective method for measuring the similarity of complex odor samples has been established in this paper. The success is attributed to the introduction of the AOFMs that can represent all the information of an odor sample and show more details of an odor sample than its original signal form. The AOFMs also provide the basis to apply a two-way similarity measurement technique, which makes the odor discrimination more objective and effective than other techniques such as PCA, SIMCA, and PARAFAC when complicated odor samples are encountered.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-8220/18/8/2658/s1>, Equations (S1)–(S6), Table S1: The SSR of the Pipe Tobacco samples, Table S2: The SSR of the Tobacco Smalls samples, Figure S1: SSR of Pipe Tobacco and Tobacco Smalls samples along with growth of components by PARAFAC and PARAFAC2.

Author Contributions: F.G. is the group leader and he is responsible for the project management and in charge of revising this manuscript. W.G. in charge of data analysis and the preparation of this manuscript. H.K. and J.W. are in charge of preparing samples and collecting data.

Funding: This work was supported by the China Tobacco Guangdong Industrial Co., Ltd. (05XM-QK[2012]019) and National Natural Science Foundation of China (20875106).

Acknowledgments: The authors thank the anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mori, K.; Nagao, H.; Yoshihara, Y. The olfactory bulb: Coding and processing of odor molecule information. *Science* **1999**, *286*, 711–715. [[CrossRef](#)] [[PubMed](#)]
2. Persaud, K.; Dodd, G. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **1982**, *299*, 352–355. [[CrossRef](#)] [[PubMed](#)]
3. Gardner, J.W.; Bartlett, P.N. Performance definition and standardization of electronic noses. *Sens. Actuators B Chem.* **1996**, *33*, 60–67. [[CrossRef](#)]
4. Sanaeifar, A.; Mohtasebi, S.S.; Ghasemi-Varnamkhasti, M.; Ahmadi, H. Application of MOS based electronic nose for the prediction of banana quality properties. *Measurement* **2016**, *82*, 105–114. [[CrossRef](#)]
5. Biondi, E.; Blasioli, S.; Galeone, A.; Spinelli, F.; Cellini, A.; Lucchese, C.; Braschi, I. Detection of potato brown rot and ring rot by electronic nose: From laboratory to real scale. *Talanta* **2014**, *129*, 422–430. [[CrossRef](#)] [[PubMed](#)]
6. Russo, M.; Di Sanzo, R.; Cefaly, V.; Cefaly, W.; Carabetta, S.; Serra, D.; Fuda, S. Non-destructive flavour evaluation of red onion (*Allium cepa* L.) ecotypes: An electronic-nose-based approach. *Food Chem.* **2013**, *141*, 896–899. [[CrossRef](#)] [[PubMed](#)]
7. Han, F.; Huang, X.; Teye, E.; Gu, F.; Gu, H. Nondestructive detection of fish freshness during its preservation by combining electronic nose and electronic tongue techniques in conjunction with chemometric analysis. *Anal. Methods* **2013**, *6*, 529–536. [[CrossRef](#)]
8. Zhang, M.; Wang, X.; Liu, Y.; Xu, X.; Zhou, G. Species discrimination among three kinds of puffer fish using an electronic nose combined with olfactory sensory evaluation. *Sensors* **2012**, *12*, 12562–12571. [[CrossRef](#)] [[PubMed](#)]
9. Tian, X.Y.; Cai, Q.; Zhang, Y.M. Rapid classification of hairtail fish and pork freshness using an electronic nose based on the PCA method. *Sensors* **2012**, *12*, 260–277. [[CrossRef](#)] [[PubMed](#)]
10. Giungato, P.; Laiola, E.; Nicolardi, V. Evaluation of Industrial Roasting Degree of Coffee Beans by Using an Electronic Nose and a Stepwise Backward Selection of Predictors. *Food Anal. Methods* **2017**, *10*, 3424–3433. [[CrossRef](#)]
11. Upadhyay, R.; Sehwal, S.; Mishra, H.N. Electronic nose guided determination of frying disposal time of sunflower oil using fuzzy logic analysis. *Food Chem.* **2017**, *221*, 379–385. [[CrossRef](#)] [[PubMed](#)]
12. Qiu, S.; Wang, J. The prediction of food additives in the fruit juice based on electronic nose with chemometrics. *Food Chem.* **2017**, *230*, 208–214. [[CrossRef](#)] [[PubMed](#)]

13. Van Geffen, W.H.; Bruins, M.; Kerstjens, H.A. Diagnosing viral and bacterial respiratory infections in acute COPD exacerbations by an electronic nose: A pilot study. *J. Breath Res.* **2016**, *10*, 036001. [[CrossRef](#)] [[PubMed](#)]
14. Brekelmans, M.P.; Niki, F.; Paul, B.; Lieuwe, D.B.; Peter, J.S.; Paul, P.T.; Daniëlle, M.G. Smelling the Diagnosis: The Electronic Nose as Diagnostic Tool in Inflammatory Arthritis. A Case-Reference Study. *PLoS ONE* **2016**, *11*, e0151715. [[CrossRef](#)] [[PubMed](#)]
15. Zakko, L.; Chan, D.K.; Visrodia, K.; Allen, J.; Lutzke, L.S.; Clemens, M.A.; Leggett, C.L.; Wang, K.K. Detection of Barrett's Esophagus via Electronic-Nose Device Analysis of Exhaled Volatile Organic Compounds in Breath Samples. *Gastroenterology* **2017**, *152*, S164. [[CrossRef](#)]
16. Lidia, E.; Laura, C.; Selena, S. Electronic Nose Testing Procedure for the Definition of Minimum Performance Requirements for Environmental Odor Monitoring. *Sensors* **2016**, *16*, 1548. [[CrossRef](#)]
17. Szulczyński, B.; Wasilewski, T.; Wojnowski, W.; Majchrzak, T.; Dymerski, T.; Namieśnik, J.; Gębicki, J. Different Ways to Apply a Measurement Instrument of E-Nose Type to Evaluate Ambient Air Quality with Respect to Odour Nuisance in a Vicinity of Municipal Processing Plants. *Sensors* **2017**, *17*, 2671. [[CrossRef](#)] [[PubMed](#)]
18. Huang, T.; Jia, P.; He, P.; Duan, S.; Yan, J.; Wang, L. A Novel Semi-Supervised Method of Electronic Nose for Indoor Pollution Detection Trained by M-S4VMs. *Sensors* **2016**, *16*, 1462. [[CrossRef](#)] [[PubMed](#)]
19. Onodera, T.; Toko, K. Towards an Electronic Dog Nose: Surface Plasmon Resonance Immunosensor for Security and Safety. *Sensors* **2014**, *14*, 16586. [[CrossRef](#)] [[PubMed](#)]
20. Lee, D.S.; Lim, J.W.; Lee, S.M.; Huh, J.S.; Lee, D.D. Fabrication and characterization of micro-gas sensor for nitrogen oxides gas detection. *Sens. Actuators B Chem.* **2000**, *64*, 31–36. [[CrossRef](#)]
21. Alizadeh, M.; Pirsá, S.; Faraji, N. Determination of Lemon Juice Adulteration by Analysis of Gas Chromatography Profile of Volatile Organic Compounds Extracted with Nano-Sized Polyester-Polyaniline Fiber. *Food Anal. Methods* **2017**, *10*, 2092–2101. [[CrossRef](#)]
22. Jeon, D.B.; Hong, Y.S.; Lee, G.H.; Park, Y.M.; Lee, C.M.; Nho, E.Y.; Choi, J.Y.; Jamila, N.; Khan, N.; Kim, K.S. Determination of volatile organic compounds, catechins, caffeine and theanine in Jukro tea at three growth stages by chromatographic and spectrometric methods. *Food Chem.* **2017**, *219*, 443–452. [[CrossRef](#)] [[PubMed](#)]
23. Skov, T.; Bro, R. A new approach for modelling sensor based data. *Sens. Actuators B Chem.* **2005**, *106*, 719–729. [[CrossRef](#)]
24. Liang, Z.; Tian, F.; Yang, S.X.; Zhang, C.; Sun, H.; Liu, T. Study on Interference Suppression Algorithms for Electronic Noses: A Review. *Sensors* **2018**, *18*, 1179. [[CrossRef](#)] [[PubMed](#)]
25. Gancarz, K.; Wawrzyniak, J.; Gawrysiak-Witulska, M.; Wiącek, D.; Nawrocka, A.; Tadla, M.; Rusinek, R. Application of electronic nose with MOS sensors to prediction of rapeseed quality. *Measurement* **2017**, *103*, 227–234. [[CrossRef](#)]
26. Liu, J.; Wang, W.X.; Yang, Y.J.; Yan, Y.N.; Wang, W.Y.; Wu, H.Z.; Ren, Z.H. A Rapid Discrimination of Authentic and Unauthentic Radix Angelicae Sinensis Growth Regions by Electronic Nose Coupled with Multivariate Statistical Analyses. *Sensors* **2014**, *14*, 20134–20148. [[CrossRef](#)] [[PubMed](#)]
27. Zhu, P.; Du, J.; Benlian, X.U.; Lu, M. Modified unsupervised discriminant projection with an electronic nose for the rapid determination of Chinese mitten crab freshness. *Anal. Methods* **2017**, *9*, 1806–1815. [[CrossRef](#)]
28. Ordukaya, E.; Karlik, B. Quality Control of Olive Oils Using Machine Learning and Electronic Nose. *J. Food Qual.* **2017**, *2017*, 1–7. [[CrossRef](#)]
29. Li, S.; Li, X.R.; Wang, G.L.; Nie, L.X.; Yang, Y.J.; Wu, H.Z.; Wei, F.; Zhang, J.; Tian, J.G.; Lin, R.C. Rapid discrimination of Chinese red ginseng and Korean ginseng using an electronic nose coupled with chemometrics. *J. Pharm. Biomed. Anal.* **2012**, *70*, 605–608. [[CrossRef](#)] [[PubMed](#)]
30. Men, H.; Fu, S.; Yang, J.; Cheng, M.; Shi, Y.; Liu, J. Comparison of SVM, RF and ELM on an Electronic Nose for the Intelligent Evaluation of Paraffin Samples. *Sensors* **2018**, *18*, 285. [[CrossRef](#)] [[PubMed](#)]
31. Padilla, M.; Montoliu, I.; Pardo, A.; Perera, A.; Marco, S. Feature extraction on three way enose signals. *Sens. Actuators B Chem.* **2006**, *116*, 145–150. [[CrossRef](#)]
32. Ballabio, D.; Consonni, V.; Todeschini, R. Classification of multiway analytical data based on MOLMAP approach. *Anal. Chim. Acta* **2007**, *605*, 134–146. [[CrossRef](#)] [[PubMed](#)]
33. Men, H.; Shi, Y.; Jiao, Y.; Gong, F.; Liu, J. Electronic nose sensors data feature mining: A synergetic strategy for the classification of beer. *Anal. Methods* **2018**, *10*, 2016–2025. [[CrossRef](#)]

34. Yan, J.; Duan, S.; Huang, T.; Huang, T.; Wang, L. Hybrid feature matrix construction and feature selection optimization-based multi-objective QPSO for electronic nose in wound infection detection. *Sens. Rev.* **2016**, *36*, 23–33. [[CrossRef](#)]
35. Wen, T.; Yan, J.; Huang, D.; Lu, K.; Deng, C.; Zeng, T.; Yu, S.; He, Z. Feature Extraction of Electronic Nose Signals Using QPSO-Based Multiple KFDA Signal Processing. *Sensors* **2018**, *18*, 388. [[CrossRef](#)] [[PubMed](#)]
36. Jian, Y.; Huang, D.; Jia, Y.; Lu, K.; Huang, Y.; Wen, T.; Zeng, T.; Zhong, S.; Xie, Q. A Novel Extreme Learning Machine Classification Model for e-Nose Application Based on the Multiple Kernel Approach. *Sensors* **2017**, *17*, 1434. [[CrossRef](#)] [[PubMed](#)]
37. Luo, H.; Jia, P.; Qiao, S.; Duan, S. Enhancing electronic nose performance based on a novel QPSO-RBM technique. *Sens. Actuators B Chem.* **2018**, *259*, 241–249. [[CrossRef](#)]
38. Guo, W.; Gan, F.; Kong, H.; Wu, J. Signal model of electronic noses with metal oxide semiconductor. *Chemom. Intell. Lab. Syst.* **2015**, *143*, 130–135. [[CrossRef](#)]
39. Bai, W.; Ren, M.; Hopke, P.K.; Gan, F. A Similarity Measure for Two-Dimensional Fluorescent Spectra. *Appl. Spectrosc.* **2009**, *63*, 810–814. [[CrossRef](#)] [[PubMed](#)]
40. Cellini, A.; Blasioli, S.; Biondi, E.; Bertaccini, A.; Braschi, I.; Spinelli, F. Potential Applications and Limitations of Electronic Nose Devices for Plant Disease Diagnosis. *Sensors* **2017**, *17*, 2596. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).