Received: 31 July 2017 | Revised: 23 April 2018 | Accepted: 5 July 2018

DOI: 10.4218/etrij.2017-0018

ORIGINAL ARTICLE

WILEY ETRI Journal

Sensor array optimization techniques for exhaled breath analysis to discriminate diabetics using an electronic nose

Jin-Young Jeon[1] | **Jang-Sik Choi**[1] | **Joon-Boo Yu**[1] | **Hae-Ryong Lee**[2] | **Byoung Kuk Jang**[3] | **Hyung-Gi Byun**[1]

[1]Division of Electronics, Information and Communication Engineering, Kangwon National University, Samcheok, Rep. of Korea.

[2]SW& Content Research Laboratory, Electronics, Information and Communication Engineering, Deajeon, Rep. of Korea.

[3]Department of Internal Medicine, Keimyung University, Deagu, Rep. of Korea.

Correspondence
Hyung-Gi Byun, Division of Electronics, Information and Communication Engineering, Kangwon National University, Samcheok, Rep. of Korea.
Email: byun@kangwon.ac.kr

Funding information
This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP), Grant/Award Number: (No. 2015-0-00318, Olfactory Bio Data Based Emotion Enhancement Interactive Content Technology Development).

Disease discrimination using an electronic nose is achieved by measuring the presence of a specific gas contained in the exhaled breath of patients. Many studies have reported the presence of acetone in the breath of diabetic patients. These studies suggest that acetone can be used as a biomarker of diabetes, enabling diagnoses to be made by measuring acetone levels in exhaled breath. In this study, we perform a chemical sensor array optimization to improve the performance of an electronic nose system using Wilks' lambda, sensor selection based on a principal component (B4), and a stepwise elimination (SE) technique to detect the presence of acetone gas in human breath. By applying five different temperatures to four sensors fabricated from different synthetic materials, a total of 20 sensing combinations are created, and three sensing combinations are selected for the sensor array using optimization techniques. The measurements and analyses of the exhaled breath using the electronic nose system together with the optimized sensor array show that diabetic patients and control groups can be easily differentiated. The results are confirmed using principal component analysis (PCA).

KEYWORDS
B4 (based on principal component), diabetes, E-nose, principal component analysis (PCA), sensor array optimization, stepwise elimination (SE), Wilks' lambda

1 | INTRODUCTION

An electronic nose is a system that measures and analyzes odors using gas sensors. Many researchers have studied various applications of the electronic nose since Persaud et al introduced this device in 1982 [1]. One of its applications is to discriminate human disease [2]. Disease discrimination using the electronic nose is carried out by measuring the amount of a specific gas contained in the exhaled breath of a patient. The specific gas exists in high concentrations in the breath of patients who have specific diseases, and the gas may be used as a biomarker to discriminate the disease [3]. For example, the conventional method used to diagnose and monitor diabetes utilizes an invasive method to collect the patient's blood, but this can be inconvenient and painful for the patient. However, in many studies, acetone was detected in the breath of diabetic patients [4,5]. These studies indicated that acetone could be used as a biomarker of diabetes; therefore, by determining the presence of acetone in the breath, diabetes

diagnoses can be made. Currently, several researchers are attempting to perform diabetes screening and discrimination studies using the human breath [6-9].

To identify acetone from exhaled gases containing various interference gases using an electronic nose, suitable sensors are required. Thus, an integrated multi sensor, such as a sensor array, can be applied to an electronic nose, and it is necessary to optimize the performance of the composite sensor array when multiple sensors are used. Therefore, in this study, a chemical sensor array was optimized for application to the electronic nose to detect low concentrations of acetone using the B4 [10,11] Wilks' lambda [12,13], and SE [14,15] techniques. The performances of the arrays employed in each technique were compared using principal component analysis (PCA) and the Euclidean distance (ED), and the best performance array was chosen as the optimal sensor array. A total of 25 samples (12 diabetic patients and 13 controls) were measured, and the measurement data were analyzed using PCA. The sensors were provided by the Korea Institute of Science and Technology (KIST).

2 | EXPERIMENTAL

2.1 | Sensor array optimization

2.1.1 | Nominated sensor arrays

In this study, the candidate sensors used for the sensor array optimization are listed in Table 1. The candidates were classified as 20 sensing combinations, depending on the four synthetic materials and the temperature of the heater.

Both SnO_2 [16,17] and WO_3 [18] have a high selectivity for acetone, which is a biomarker of diabetes, and can measure the interference gases present in human breath. Therefore, they were chosen as the experimental sensors in this study. In addition, the experiments were performed at various temperatures (245°C, 285°C, 310°C, 325°C, and 340°C) to determine the optimum operating temperature of each sensor.

The sensor measured the target and interference gases at five different operating temperatures. For the same sensor, the sensing range varied depending on the heating temperatures. Thus, even if the sensors contained the same synthetic material, when the operating temperature changed, it was treated as a separate sensor. In Table 1, S1 through S4 indicate each of the sensors.

TABLE 1 Nominated sensors for the sensor array optimization

Sensor	Synthetic stuff	Heater temperature (°C)
S1	Au/N-SnO_2	245, 285, 310, 325, 340
S2	Au/N-WO_3	245, 285, 310, 325, 340
S3	N-WO_3	245, 285, 310, 325, 340
S4	N-SnO_2	245, 285, 310, 325, 340

2.1.2 | Gas measurement

Figures 1 and 2 show the measurement system in which the candidate sensors were installed and a block diagram of the system, respectively. The system was designed for this study. Table 2 lists the gases used for the measurement, which are given by G1 through G5.

Acetone (G1) was the target gas for the identification of diabetes, and ethanol (G2), xylene (G4), toluene (G5) [19], and NO (G3) [20] were selected as the interference gases contained in the human breath.

All gases were measured at a flow rate of 200 mL/m after stabilizing the sensor array using air for 12 hours. The sensor array stabilization was performed once after turning on the measurement system.

When measuring the gas, the injection time was 300 seconds, and the sensor recovery time was 600 seconds. All gases were measured under the same conditions, and the data were sampled every second.

Figure 3 shows the result of the measured G1 by S1 with a heater temperature of 285°C, and the *x*-axis represents the time while the *y*-axis represents the sensor resistance.

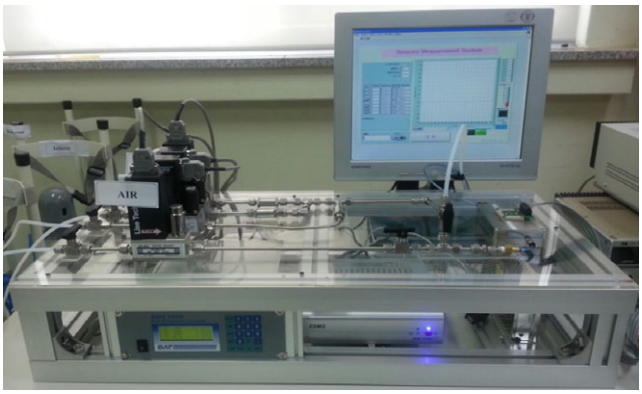

FIGURE 1 Gas measurement system used in the sensor array optimization experiment

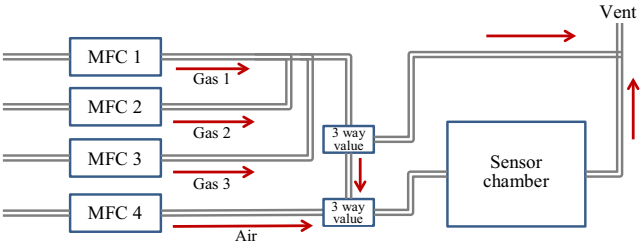

FIGURE 2 Block diagram of the gas measurement system

TABLE 2 Target gas and the interference gases used in the sensor array optimization experiment

Gas number	Gas (ppm)	Note
G1	Acetone 10	Target gas
G2	Ethanol 5	Interference gas
G3	NO 10	Interference gas
G4	Xylene 10	Interference gas
G5	Toluene 10.5	Interference gas

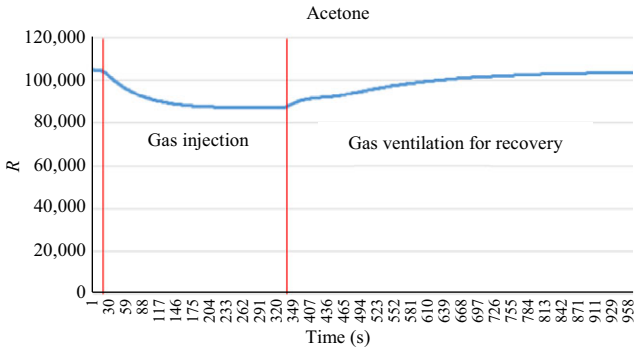

FIGURE 3 Raw response graph of acetone measured by S1_285

The red bars in Figure 3 show the events that occurred in each section. The gas injection section in Figure 3 shows the changes of the sensor response curve during the gas measurement, and the second section shows the response curve during the recovery of the sensor array after the gas measurement.

2.1.3 | Feature selection and normalization of data

For the sensor array optimization, the data were used after converting to the sensitivity value using (1).

$$R_s[\%] = \frac{R_{gas} - R_{air}}{R_{air}} \times 100. \quad (1)$$

In (1), R_{gas} is the response value of the sensor and R_{air} is the value when it is in the steady state before being exposed to the gas. Figure 4 shows a graph of the converted sensitivity value, and the red line value in this figure was selected as a feature of the measurement data. It was applied for all of the measurement data.

In addition, the patterns were normalized using (2) to compare the results of each sensor measurement on the same scale.

$$x'_{ij} = \frac{x_{ij}}{\sum_{j=0}^{N} x_{ij}} \times 100. \quad (2)$$

In (2), i is the sequence number of the test gases and j is the sequence number of the candidate sensors. In addition, x'_{ij} is the normalized pattern and x_{ij} is the sensitivity value of the sensor to the gas. The parameter N represents the number of sensors.

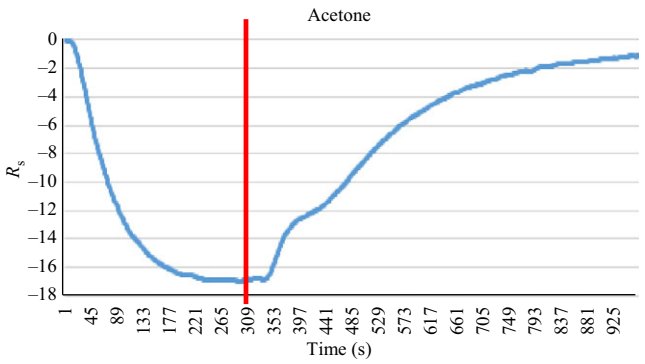

FIGURE 4 Sensitivity value graph converted of the raw response of acetone measured by S1_285

TABLE 3 Normalized data set of the values measured with each sensor

Num.	Sensor	G1	G2	G3	G4	G5
1	S1_245	6.151	6.199	4.184	4.539	4.369
2	S2_245	5.280	5.042	13.237	4.934	2.014
3	S3_245	3.355	3.279	8.918	2.780	1.697
4	S4_245	4.012	4.227	3.751	3.297	3.539
5	S1_285	5.506	5.051	3.312	5.377	6.678
6	S2_285	5.137	5.246	13.328	5.178	3.293
7	S3_285	3.349	3.424	8.645	2.943	3.026
8	S4_285	3.761	3.505	1.339	3.203	4.906
9	S1_310	3.281	4.240	2.162	4.695	5.683
10	S2_310	3.868	4.467	5.959	4.261	2.429
11	S3_310	2.343	2.942	4.412	2.659	2.086
12	S4_310	2.379	3.123	0.656	3.382	4.215
13	S1_325	5.602	4.847	0.967	5.645	3.837
14	S2_325	6.575	6.261	8.528	7.570	4.847
15	S3_325	4.504	4.313	6.255	5.108	4.545
16	S4_325	4.315	3.532	0.344	4.158	3.236
17	S1_340	8.000	7.980	0.657	7.088	9.669
18	S2_340	9.593	9.301	6.875	10.997	11.023
19	S3_340	6.978	7.248	6.101	7.328	10.353
20	S4_340	6.013	5.773	0.372	4.859	8.554

Table 3 lists the normalized data acquired from all the sensors, and S1_250 in the sensor column refers to the sensor and the operating temperature.

2.1.4 | Sensor array optimization technique

Wilks' lambda

Wilks' lambda is a statistical technique that analyzes the difference between the response value of each sensor

(group) and the average value of each sensor response, and has a value between 0 and 1.

Wilks' lambda is defined as follows [13]:

$$\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|W+B|}. \tag{3}$$

The parameter T represents the degree of spread of the data vectors from the mean vector of the entire sample. Thus, it refers to the variance of the entire input data. The parameter B is a matrix indicating how the mean vector of each group is spread out from the overall mean vector, and represents the variance among the groups. The group included sensors, and if B (the variance between the groups) is large, then each sensor would respond differently to the same gas.

The parameter B is defined as follows:

$$B = \sum_{i=1}^{g} n_i (\overline{X}_i - \overline{X})(\overline{X}_i - \overline{X})'. \tag{4}$$

The parameter g is the number of groups (sensors) and n_i is the number of patterns that belong to the i-th group. The parameter $\bar{x}_i$ is the mean of the i-th group and $\bar{x}$ is the overall mean. The parameter W is a matrix indicating how the data vectors of each group are spread out from the average vector of the group and represents the variance within the group. If W (the variance within the groups) is large, the sensor would respond differently for various gases.

Therefore, the larger the W, the better the selectivity of the sensor.

The parameter W is defined by (5).

$$W = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)' \tag{5}$$

The parameter x_{ij} is the value of the j-th sensor of the i-th group.

The parameter $\Lambda(m)$ is expressed as follows:

$$\Lambda(m) = \frac{|W|}{|T|} = \frac{d_{11}^{(0)}}{t_{11}^{(0)}} \frac{d_{22}^{(0)}}{t_{22}^{(0)}} \cdots \frac{d_{mm}^{(m-1)}}{t_{mm}^{(m-1)}}. \tag{6}$$

Then, the right-hand term of (6) is separated and shown as $U_{(m)}$, such as (7) for convenience, where m is the sensor. The parameter $U_{(m)}$ is a value obtained by dividing the variance within the group of each sensor by the total variance of each sensor, and the closer the value is to 1, the greater the discriminative power of the sensor.

$$U_{(m)} = \frac{d_{mm}^{(m-1)}}{t_{mm}^{(m-1)}}. \tag{7}$$

B4 (technique based on principal components)

The sensor selection based on the principal component is a technique that uses eigenvalues and eigenvectors, which are the PCA results of the input data. The eigenvector indicates the direction of the force acting on a certain matrix, and the eigenvalue represents the correlation coefficient (scale) of the eigenvector. To obtain the eigenvalues and eigenvectors, a covariance matrix of the input data is required.

Then, the eigenvalues and eigenvectors of the input data can be obtained from the covariance matrix. The covariance matrix is a square matrix filled with the variance of each of the X and Y variables as well as a covariance matrix between X and Y and Y and X, when the input data X and Y exist.

Figure 5 shows an example of a 2×2 covariance matrix.

Equations (8) and (9) are used to obtain the variance and covariance, and n is the number of observations of the variable. The parameter X_i is the i-th observation of the X variable and $\bar{X}$ represents the average of all observations. The parameter Y_i is the i-th observation value of the Y variable and $\bar{Y}$ is the average of all observation values of the Y variable.

$$\mathrm{var}(X) = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})}{(n-1)}, \tag{8}$$

$$\mathrm{cov}(X) = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)}. \tag{9}$$

Thus, because the covariance matrix is a matrix composed of the variance and covariance values of the input data, the eigenvector of this matrix represents the direction in which the input data are dispersed.

If the eigenvectors are arranged from the largest eigenvalues, the principal components will be obtained in order of importance. Then, if this is applied to the sensor selection, the input data are the response values measured by the candidate sensors. The eigenvectors and the eigenvalues obtained from these input data represent the direction and scale of the covariance matrix of the data measured by each sensor [21]. The larger the eigenvalue and eigenvector, the greater will be the variance of the gas measurement values of the sensor, and this is related to the selectivity of the sensor.

In this study, the optimal sensors were chosen using the B4 technique, which is one of the sensor-selection techniques based on the principal component. The B4 technique selects candidate sensors having the largest absolute value from among the eigenvectors corresponding to eigenvalues greater than a predefined constant value ($\lambda_0 = 0.7$) [10].

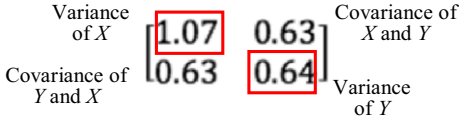

FIGURE 5 Example of 2×2 covariance matrix

Thus, B4 is a technique that is used to select only a sensor having a variance greater than or equal to the threshold value from the principal component of the input data.

Stepwise elimination (SE)

By applying a leave-one-out method to various candidate sensors, the sensor selection based on SE is a technique that eliminates the sensors that have the greatest effect on the selectivity and sensitivity of the sensor array. A sensor is removed repeatedly until there is no more improvement in the sensitivity and selectivity, and finally, the remaining sensors are selected.

For example, when there are four candidate sensors a, b, c, and d, the steps in the sensor-selection process using the SE technique are as listed in Table 4.

The sensitivity (root sum square [RSS]) of each candidate sensor was calculated according to (10).

$$\mathrm{RSS} = \left(\sum_{c=1}^{n} X_c^2 \right)^{1/2}. \tag{10}$$

The parameter n is the number of gas or the number of patterns and X_c is the response value of the c-th sensor. In addition, the sensitivity (sum of the root sum square [SRSS]) of the sensor array was obtained by (11).

The sensitivity of the sensor arrays represented the sum of the sensitivity of each candidate sensor.

$$\mathrm{SRSS} = \sum_{i=1}^{m} \left(\sum_{c=1}^{n} X_c^2 \right)^{1/2}. \tag{11}$$

The parameter M is the number of sensors included in the array for which the sensitivity is to be calculated. Then, the sensor array selectivity (sum of the Euclidean distance [SED]) was obtained by (12) [14].

$$\mathrm{SED} = \sum_{k=1}^{N-1} \left(\sum_{l=k+1}^{N} d_{\mathrm{kl}} \right), \tag{12}$$

$$\text{where } d_{\mathrm{kl}} = \sqrt{\sum_{j=1}^{m} \left(x_{\mathrm{kj}} - x_{\mathrm{lj}} \right)^2}.$$

The parameter N is the number of gases and m is the number of sensors. The parameter X_j is the response value of the sensor j corresponding to each gas k and l. The parameter d_{kl} is the ED between the gases k and l, and the selectivity of the sensor array is the sum of the ED between each gas pattern. Thus, when the sensitivity and selectivity of a sensor array increase, the sensor array shows a greater sensitivity to the measurement gases, and the discrimination between the gases was smooth.

TABLE 4 Sensor-selection procedure using the SE technique

Step	Procedure
1	Perform the selectivity and sensitivity calculation of the array consisting of b, c, and d, but not sensor a
2	Perform the selectivity and sensitivity calculation of the array consisting of a, c, and d, but not sensor b
3	Perform the selectivity and sensitivity calculation of the array consisting of a, b, and d, but not sensor c
4	Perform the selectivity and sensitivity calculation of the array consisting of a, b, and c, but not sensor d
5	Select the array with the highest selectivity and sensitivity values from among the array configurations of steps 1, 2, 3, and 4
6	Repeat steps 1 through 5 with the selected sensors in step 5 until there is no more improvement in the sensitivity and selectivity

2.2 | Breath measurement for diabetes discrimination

2.2.1 | Breath measurement system

Figure 6 shows a block diagram of the measurement system designed to measure the breath of the diabetic and control samples by applying the sensors selected using the sensor array optimization. The breath of the patient and the control samples were collected by a Tedlar bag and the collected breath gases were sampled by a solid phase microextraction (SPME) fiber (75-μm CAR/PDMS) [22–24]. Then, the sampled SPME fiber was inserted into the system to be measured.

Figure 7 shows the developed electronic nose system for the breath measurement and the connected PC employed to display and store the measurement signal in real time.

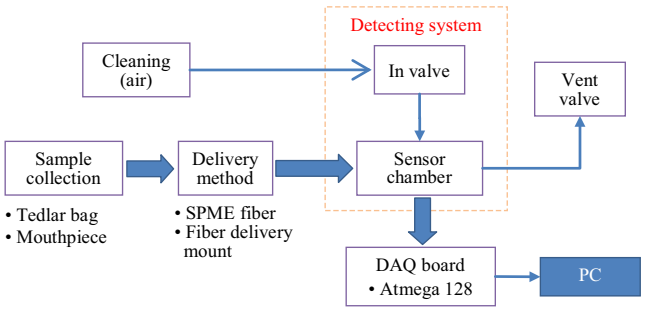

FIGURE 6 Block diagram of the breath measurement system

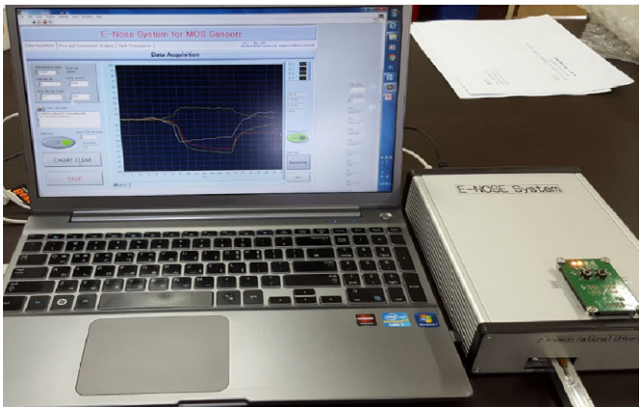

FIGURE 7 Electronic nose system for breath measurement

2.2.2 | Breath measurement and feature extraction

For breath measurements, first, approximately 500-mL exhalations of the diabetic and control samples were collected using a Tedlar bag. All of the samples were collected 2 hours after breakfast and after brushing teeth with bottled water (without toothpaste) in the same location of the hospital.

Then, to sample the breath, the SPME fiber was inserted into the Tedlar bag for 20 minutes. Finally, the sampled SPME fiber was inserted into the measurement system and measured for 10 minutes (needle open for 5 minutes and recovery for 5 minutes).

Figure 8 shows the sensor responses obtained for the control sample measured by this procedure. The acquired data from the sensor array were converted to Rs [%] using (1), and were stored.

The red bars in Figure 8 show the events that occurred in each section. The fiber insertion and waiting section in Figure 8 shows the wait period of the sampled SPME fiber after being ready in the measurement system. In addition, the needle open section is the time required to measure the sampled breath gas by opening the needle of the inserted SPME fiber. The feature-extraction section is selected as the feature to analyze the measured breath, and the final pumping for the recovery section is the period during which the motor pump is operated to recover the sensor array.

TABLE 5 Results of the Wilks' lambda calculation

Num.	Sensor	U (m)
20	**S4_340**	**0.998171**
15	**S3_325**	**0.993984**
1	**S1_245**	**0.990333**
5	S1_285	0.971844
2	S2_245	0.921099

2.2.3 | Cluster trend analysis

To analyze the cluster trends of the sensor array patterns selected by each technique and the measured breath pattern, the PCA technique was used.

PCA is a technique that projects high-dimensional input data to a low dimension using the eigenvalues and eigenvectors of the input data, and is similar to the B4 technique, which is a sensor-selection technique based on the principal component. Thus, by showing high-dimensional input data on a two-dimensional plane, the cluster tendency of the data can be analyzed.

3 | RESULTS AND DISCUSSIONS

3.1 | Sensor array optimization

3.1.1 | Wilks' lambda

The Wilks' Lambda technique was applied to the response data normalized by the candidate sensors and Table 5 lists

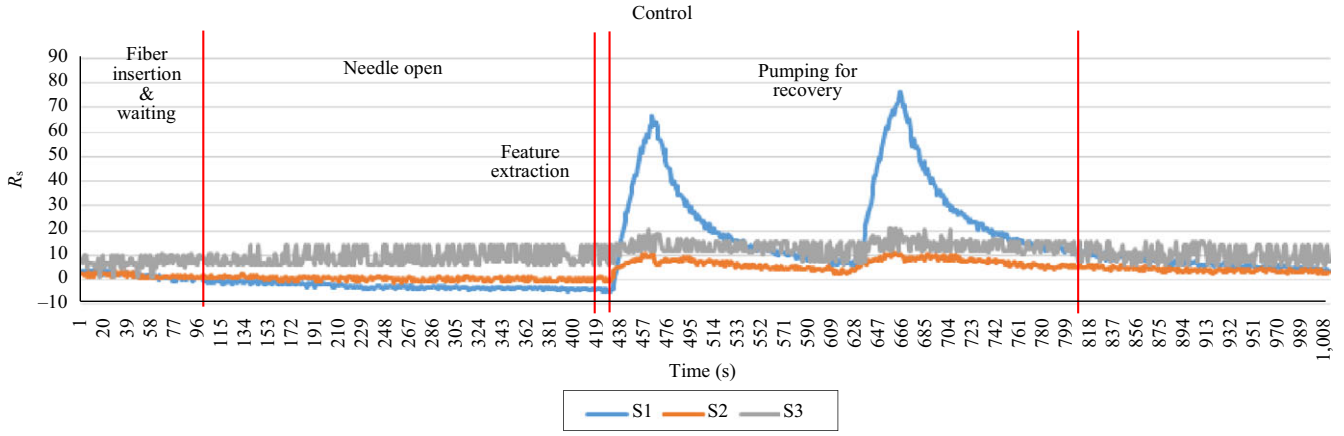

FIGURE 8 Sensors response of breath sample of control group

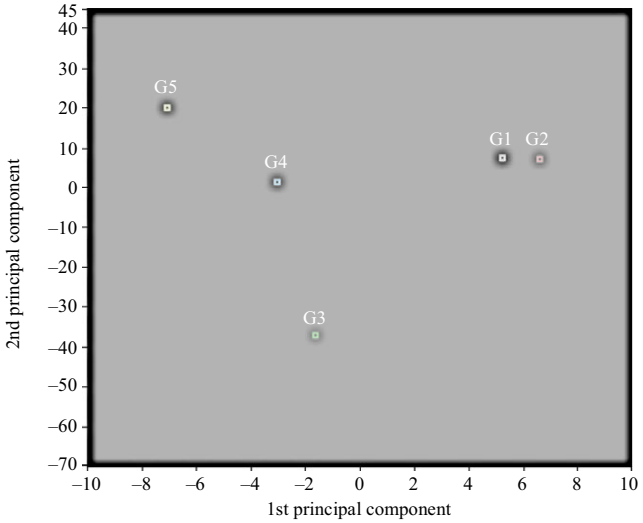

FIGURE 9 Principal component analysis result of sensor array selected by Wilks lambda

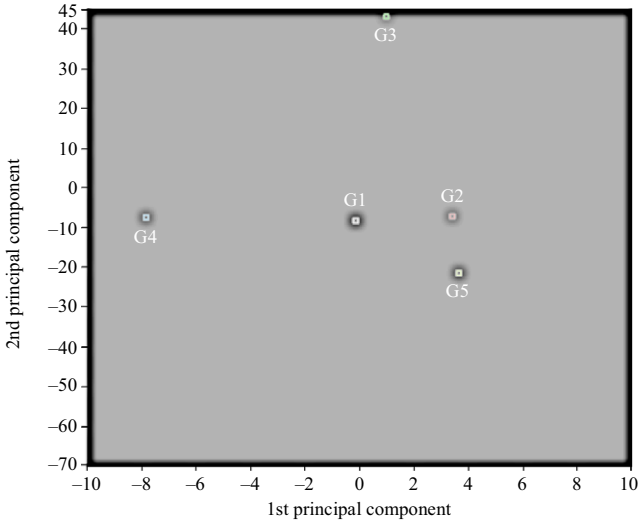

FIGURE 10 Principal component analysis result of sensor array selected by B4

TABLE 6 Results of the B4 technique performed for sensor selection

Eigen values		Num. Sensor	1 S1_245	2 S2_245	13 S1_325	Abs. max
Eigen values 1	85.444	Eigen vectors.1	0.036	**−0.454**	0.167	0.454
Eigen values 2	5.200	Eigen vectors.2	−0.284	−0.089	**−0.505**	0.505
Eigen values 3	1.958	Eigen vectors.3	**−0.480**	−0.068	0.112	0.480

the results. When the value of $U(m)$ was large, the discriminative power of the sensor was high. Therefore, Table 5 lists the sensors in descending order based on $U(m)$.

An array using the upper three sensors was composed (S4_340, S3_325, and S1_245) based on $U(m)$ in Table 5. Figure 9 shows the PCA result of the data measured by the three sensors that were selected by the Wilks' lambda technique.

As shown in Figure 9, the target gas G1, interference gases, and respective interference gases were distinguishable. However, the interference gas G2, which was similar to the target gas G1, was located close to G1.

3.1.2 | B4 (technique based on principal components)

Table 6 lists the applied result of the B4 technique to the response data normalized by the candidate sensors. Three sensors (S1_245, S2_245, and S1_325) were selected.

Figure 10 shows the PCA result of the data measured by the three sensors that were selected by the B4 technique. The discrimination between target gas G1 and the interfering gases was apparent. When comparing Figure 9, while the distance between G1 and G2 was large, the distance between G2 and G5 was small.

3.1.3 | Stepwise elimination

Figure 11 shows the sensor-selection process results obtained using the SE technique. As a result of applying the SE technique, 17 sensors were removed, and sensors 2, 16, and 20 (S2_245, S4_325, S4_340) were selected.

Figure 12 shows the PCA result obtained for the data measured by the three sensors that were selected by the SE technique. When comparing the results with Figure 10, the distance between G1 and G2 was similar, and the distance between G1 and G4 was smaller. However, there discrimination was obtained.

3.1.4 | Optimal sensor array selection

When comparing the PCA results of each technique, the discrimination power of the SE technique was the best. However, because this result was not objective, the ED of the data measured in the array obtained by each technique was calculated. Table 6 lists the calculated EDs.

From Table 7, the SE technique had the largest ED in the greatest number of items. Therefore, the array composed of S2_245, S4_325, and S4_340 sensors selected by the SE technique was the optimal sensor array.

Iteration	Sensor number																			SED	SRSS	Average
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	121.377	239.524	180.451
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20		129.378	240.754	185.066
3	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	19	20			138.262	241.817	190.039
4	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	20				148.185	243.282	195.733
5	1	2	3	4	5	6	7	8	9	10	11	12	13	16	17	20					158.004	244.125	201.065
6	1	2	3	4	5	6	8	9	10	11	12	13	16	17	20						168.353	245.446	206.899
7	1	2	3	4	5	6	8	9	11	12	13	16	17	20							179.661	246.520	213.091
8	1	2	3	4	6	8	9	11	12	13	16	17	20								194.281	247.806	221.043
9	1	2	3	4	8	9	11	12	13	16	17	20									212.345	250.156	231.250
10	2	3	4	8	9	11	12	13	16	17	20										232.848	252.883	242.866
11	2	3	8	9	11	12	13	16	17	20											262.949	253.206	258.078
12	2	8	9	11	12	13	16	17	20												301.320	255.250	278.285
13	2	8	9	12	13	16	17	20													334.605	259.019	296.812
14	2	8	12	13	16	17	20														365.122	261.881	313.501
15	2	12	13	16	17	20															385.759	262.120	323.940
16	2	12	16	17	20																405.318	262.507	333.913
17	2	16	17	20																	435.104	256.728	345.916

Discarded sensor

Selected sensors

FIGURE 11 Results of SE technique performed for sensor selection

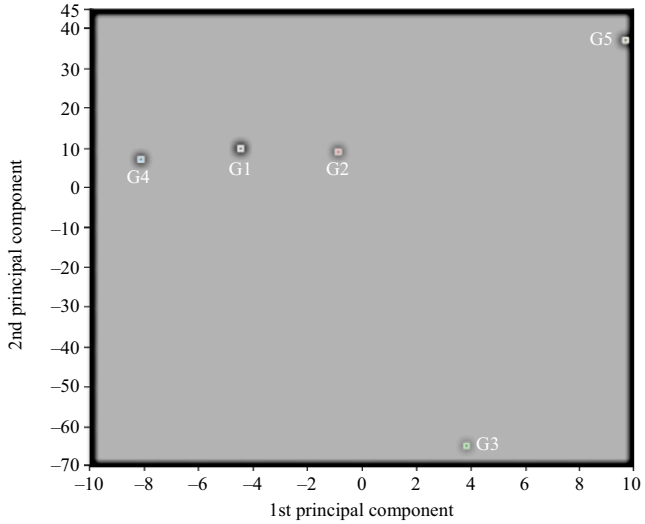

FIGURE 12 Principal component analysis result of sensor array selected by SE

TABLE 7 ED comparison of the arrays selected by the respective techniques (SE, B4, and Wilks' lambda)

	SE	B4	Wilks' lambda	All sensors
G1-G2	**3.721**	3.690	1.425	2.032
G1-G3	**75.138**	51.207	44.938	18.839
G1-G4	4.546	7.719	**10.261**	3.599
G1-G5	**30.615**	13.895	17.573	8.015
G2-G3	**73.953**	50.194	44.827	18.505
G2-G4	7.496	**11.232**	11.221	3.550
G2-G5	**29.934**	14.426	18.803	7.655
G3-G4	**72.954**	51.290	38.359	18.763
G3-G5	**101.985**	64.615	57.222	24.371
G4-G5	**34.762**	18.139	19.061	8.038
G1 to others	**114.020**	76.512	74.197	32.484

3.2 | Breath measurement and analysis

The selected optimal sensor array was installed in the electronic nose system for breath measurement to confirm the discrimination performance, and 25 breath samples (12 type-2 diabetic patients and 13 controls, IRB File No. DSMC2016-11-021-001) were measured by the system. The measured data were analyzed by PCA.

Tables 8 and 9 list the control groups and diabetic patient groups used in the study, respectively.

Figure 13 shows the results of the PCA analysis of the measured breath samples. From the result of the cluster analysis obtained by the PCA, it was possible to discriminate between the control group (green) and the patient group (red).

In this study, unlike previously published studies [2,6,7], an SPME fiber was used to sample the exhaled breath collected in a Tedlar bag. This was done in order to reduce the effects of the humidity contained in the exhaled breath, and to measure only the available gases. In addition, sensor array optimization was performed to determine the array combination with the fewest number of sensors and the highest selectivity for acetone, which is the biomarker for diabetes. Thus, compared with the results obtained in published studies [2,6], although fewer sensors were used, similar or slightly better discrimination results were obtained.

However, to obtain more reliable results, a larger number of new samples are required, and additional information (co-morbidities, treatment, etc.) pertaining to the samples should be added. In addition, it is necessary to perform repeated experiments to prove the reproducibility of the experiment, and downsizing should be performed to increase the portability of the electronic nose.

4 | CONCLUSIONS

In this study, a sensor array that could be applied to an electronic nose was optimized to discriminate the breath of diabetic samples from control samples. For sensor array

TABLE 8 Sample list of the control group for breath measurement

Sample no.	Sex	Age	Glucose (mg/dL)	BMI	Smoking	Drinking	HbA1c (%)
1	M	37	92	26.8	×	×	5
2	F	32	97	20.6	×	×	5.3
3	M	29	87	21.8	×	×	4.9
4	M	31	85	20.5	×	○	5.3
5	F	26	94	19.1	×	×	5.4
6	F	34	99	19.3	×	×	5.3
7	M	30	93	27.3	×	○	5.3
8	F	30	90	18.7	×	×	5.1
9	F	24	82	N/A	N/A	N/A	4.9
10	F	22	83	N/A	N/A	N/A	4.6
11	F	54	100	N/A	N/A	N/A	5.6
12	F	47	100	22.2	×	○	5.2
13	F	58	97	20.8	×	×	5.3
Mean	N/A	N/A	92.2	21.7	N/A	N/A	5.2
Standard deviation	N/A	N/A	6.4	3.0	N/A	N/A	0.3

TABLE 9 Sample list of the diabetic group for breath measurement

Sample no.	Sex	Age	Glucose (mg/dL)	BMI	Smoking	Drinking	HbA1c (%)
1	F	24	125	14.4	×	×	7.5
2	M	33	130	25.4	×	×	6.1
3	F	56	151	27.5	×	×	7.6
4	M	58	164	26.7	×	×	6.9
5	M	69	154	23.7	○	○	6.7
6	F	45	306	24.6	×	○	10.7
7	F	63	111	25.4	×	×	7.5
8	M	75	150	23.1	×	×	6.7
9	M	53	112	28.7	N/A	N/A	5.5
10	F	54	129	23.5	N/A	N/A	6.5
11	F	80	148	25.4	×	×	10.7
12	M	N/A	108	N/A	N/A	N/A	6.1
Mean	N/A	N/A	141.3	24.0	N/A	N/A	7.2
Standard deviation	N/A	N/A	52.9	3.7	N/A	N/A	1.7

optimization, the Wilks' lambda, B4, and SE techniques were applied, and 20 sensing combinations were used. Three sensor arrays were obtained using each technique, and the performance of each sensor array was compared using ED and PCA.

From Table 7, the performance of the sensor arrays obtained by each technique was better than that of the array that consisted of 20 sensing combinations. The three sensing combinations obtained by the SE technique showed the best performance and were selected as the optimum sensor array.

The selected sensor array was installed in the breath measurement system, and 25 samples (12 type-2 diabetic patients and 13 control subjects) were measured using the proposed system. The measured breath data were analyzed by the PCA. Based on the PCA, it was possible to discriminate between the control group (green) and the patient group (red).

This result was obtained by a combination of the sampling method using the SPME fiber (to minimize the humidity of the exhalation sample and measure only the effective gases) and the electronic nose with the optimized sensor array, which had the highest selectivity for acetone, and which contained the fewest sensors.

By performing this study, the possibility of diabetic disease screening with electronic noses was confirmed. In addition, if the sensor array optimization technique and the

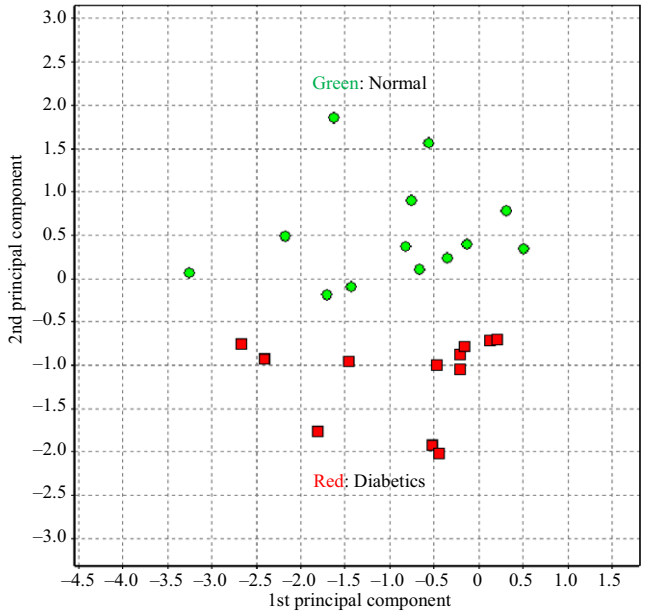

FIGURE 13 Principal component analysis result of measured breath samples

SPME fiber are applied to the electronic nose used in published studies, the performance of the electronic nose could be improved.

ACKNOWLEDGMENTS

This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2015-0-00318, Olfactory Bio Data based Emotion Enhancement Interactive Content Technology Development).

ORCID

Jin-Young Jeon http://orcid.org/0000-0002-3500-8956
Jang-Sik Choi http://orcid.org/0000-0003-2735-2287
Joon-Boo Yu http://orcid.org/0000-0001-6731-4205
Hyung-Gi Byun http://orcid.org/0000-0003-3773-3294

REFERENCES

1. K. Persaud and D. Dodd, *Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose*, Nature **299** (1982), 352–355.
2. D. Guo et al., *A novel breath analysis system based on electronic olfaction*, IEEE Trans. Biomed. Eng. **57** (2010), no. 11, 2753–2763.
3. Biomarkers Definitions Working Group, *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework*, Clin. Pharmacol. Ther. **69** (2001), no. 3, 89–95.
4. Z. Wang and C. Wang, *Is breath acetone a biomarker of diabetes? A historical review on breath acetone measurements*, J. Breath Res. **7** (2013), no. 3, pp. 037109:1–037109:18.
5. C. Deng et al., *Determination of acetone in human breath by gas chromatography–mass spectrometry and solid-phase microextraction with on-fiber derivatization*, J. Chromatogr. B **810** (2004), no. 2, 269–275.
6. J.-B. Yu et al., *Analysis of diabetic patient's breath with conducting polymer sensor array*, Sens. Actuators B Chem. 108 (2005), no. 1–2, 305–308.
7. K. Yan et al., *Design of a breath analysis system for diabetes screening and blood glucose level prediction*, IEEE Trans. Biomed. Eng. **61** (2014), no. 11, 2787–2795.
8. P. Wang et al., *A novel method for diabetes diagnosis based on electronic nose*, Biosens. Bioelectron. **12** (1997), no. 9–10, 1031–1036.
9. E. I. Mohamed et al., *Predicting type 2 diabetes using an electronic nose-based artificial neural network analysis*, Diabetes Nutr. Metab. **15** (2002), no. 4, 215–221.
10. I. T. Jolliffe, *Discarding variables in a principal component analysis. I: artificial data*, J.R. Stat. Soc. **21** (1972), no. 2, 160–173.
11. J.-S. Choi, J. Y. Jeon, and H. G. Byun, *Investigation of chemical sensor array optimization method for DADSS*, J. Sens. Sci. Technol. **25** (2016), no. 1, 13–19.
12. J.-Y. Jeon et al., *Chemical sensors array optimization based on Wilks lambda technique*, J. Sens. Sci. Technol. **23** (2014), no. 5, 299–304.
13. Y. Yin et al., *A sensor array optimization method of electronic nose based on elimination transform of Wilks statistic for discrimination of three kinds of vinegars*, J. Food Eng. **127** (2014), 43–48.
14. A. Chaudry, T. M. Hawkins, and P. J. Travers, *A method for selecting an optimum sensor array*, Sens. Actuators B Chem. **69** (2000), no. 3, 236–242.
15. H.-J. Lim et al., *A step-wise elimination method based on Euclidean distance for performance optimization regarding to chemical sensor array*, J. Sens. Sci. Technol. **24** (2015), no. 4, 258–263.
16. Y.-G. Song et al., *Metal oxide nanocolumns for extremely sensitive gas sensors*, J. Sens. Sci. Technol. **25** (2016), no. 3, 184–188.
17. Y.-S. Shim et al., *Utilization of both-side metal decoration in close-packed SnO_2 nanodome arrays for ultrasensitive gas sensing*, Sens. Actuators B Chem. **213** (2015), 314–321.
18. Y.-S. Shim et al., *Highly sensitive and selective H_2 and NO_2 gas sensors based on surface-decorated WO_3 nanoigloos*, Sens. Actuators B Chem. **198** (2014) 294–301.
19. G. Peng et al., *Diagnosing lung cancer in exhaled breath using gold nanoparticles*, Nat. Nanotechnol. **4** (2009) 669–673.
20. W. Dillon et al., *Origins of breath nitric oxide in humans*, Chest **110** (1996), no. 4, 930–938.
21. A. Gibson et al., *Deeplearning4j, A Beginner's Guide to Eigenvectors*, PCA, Covariance and Entropy, 2016, accessed July 18, 2017, available at https://deeplearning4j.org/eigenvector.
22. D. Francesco et al., *Breath analysis: trends in techniques and clinical applications*, Microchem. J. **79** (2005), no. 1–2, 405–410.
23. SUPELCO, Sigma-Aldrich Co., *Solid Phase Microextraction Fiber Assemblies*, 1999, accessed November 22, 2017, available at https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/General_Information/1/t794123.pdf.
24. SUPELCO, Sigma-Aldrich Co., *Selection Guide for Supelco SPME Fibers*, 2018, available at accessed April 12, 2018, https://www.sigmaaldrich.com/technical-documents/articles/analytical/selecting-spme-fibers.html

AUTHOR BIOGRAPHIES

Jin-Young Jeon received his MS and PhD degrees in information and communication engineering from Kangwon National University, Rep. of Korea, in 2017. Since 2012, he has been with the Division of Electronics, Information, and Communication Engineering, Kangwon National University, where he works in the Smart Sensor System Laboratory. His main research interests are data analysis and signal processing for E-Nose.

Jang-Sik Choi received his MS degree in information and communication Engineering from Kangwon National University, Rep. of Korea, in 2011. Since 2013, he has been with the Division of Electronics, Information, and Communication Engineering, where he works in the Smart Sensory System Laboratory. His research activities are focused on E-Nose system, Nanoinformatics, and Bioinformatics.

Joon-Boo Yu received his BS and MS degrees in electronics engineering from Kwandong University, Rep. of Korea, in 1992 and 1996, respectively. He received his PhD degree in Material Science and Metallurgy, Kyungpook National University, Daegu, Rep. of Korea, in 2010. He studied sensor materials and electronic nose systems at the Kyungpook National University until 2014, and since then, he has worked as a postdoctoral researcher at the Kangwon National University, studying electronic nose systems and recognition algorithms.

Hae-Ryong Lee received his BS and MS degrees in robotics from the School of Engineering and Technology, Southern Illinois University, Carbondale, the United States of America, in 1988 and 1992, respectively. He received his PhD degree in computer engineering from the Engineering Department, Chungnam National University, Daejeon, Rep. of Korea, in 2005. From 1993 to 2018, he has worked for the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, where he is now a principal researcher. His main research interests are highly sensitive and selective chemiresistive gas sensors and five sensory content designs.

Byoung Kuk Jang received his MD degree from Keimyung University School of Medicine, Daegu, Rep. of Korea, in 1995, and his PhD degree in nuclear medicine from Kyungpook National University, Daegu, Rep. of Korea, in 2007. Since 2003, he has been with the Department of Internal Medicine, Keimyung University, Daegu, Rep. of Korea, where he is now a professor. His main research interests are basic and clinical studies related to various liver and metabolic diseases.

Hyung-Gi Byun received his PhD n instrumentation and analytical science from the University of Manchester, UK, in 1995. Since 1996, he has been with the Division of Electronics, Information, and Communication Engineering, Kangwon National University (Samcheok Campus), Rep. of Korea, where he is now a full professor. His main research interests are data analysis and signal processing for intelligent sensing (E-Nose) and actuating (Olfactory Display) systems for medical and entertainment applications.