A Novel Breath Analysis System for Diabetes Diagnosis

Ke Yan
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
xjed09@gmail.com

David Zhang
Department of Computing
Hong Kong Polytechnic University
Hong Kong, China
csdzhang@comp.polyu.edu.hk

***Abstract*—A noninvasive breath analysis system for diabetes diagnosis is proposed in this paper. It utilizes commercial chemical sensors to detect acetone in human breath. The device is portable, fast and easy to operate. Experiments with real breath samples from both inpatient and outpatient diabetics validated the accuracy of this system. An optimal sensor array for diabetes diagnosis is decided according to the classification accuracy. We hope this system could contribute a step toward a practical system that could be used for diabetes screening.**

Keywords—diabetes diagnosis, breath analysis, e-nose, sensor selection

I. INTRODUCTION

Diabetes is a group of metabolic diseases featured with high blood glucose. It can cause complications related to many organs and has become a great threat to human health. With the improvement of life quality, the incidence of diabetes has been increasing. Statistics shows that about 10% people in Hong Kong are diabetics nowadays. For each patient, being diagnosed in time helps a lot for the control of diabetes. Typically, one should go to the hospital and take a blood test for diabetes diagnosis. This method is inconvenient and painful, thus not beneficial for a timely diagnosis.

Diabetes occurs when the body cannot make use of the glucose in blood. As a result, the cells have to use fat as an energy source. In the process of metabolizing fat for energy, one of the by-products is ketone body. When ketone bodies accumulate in blood, there will be an elevation of acetone concentration in the patient's breath. The relationship between diabetes and breath acetone have been confirmed by many researchers [1-3]. So we can expect to design a system which diagnoses diabetes by means of detecting the acetone concentration in human breath.

The measurement of components in breath is usually performed by gas chromatography combined with mass spectrometry (GC/MS). Despite its high sensitivity, the GC/MS technique has considerable drawbacks: it is expensive, time-consuming, not portable and difficult to handle [4]. An alternative method is using electronic noses (e-noses). E-noses use chemical sensors to sense the type and concentration of components in breath. Compared with GC/MS, e-noses are cheaper, faster, more portable and easier to operate. Recently, e-noses have gradually been used in medicine for the diagnosis of various diseases, such as diabetes, renal disease and airway inflammation [5-7].

In this paper, we propose a novel breath analysis system based on the principle of e-noses. It includes a device containing a sensor array specially chosen for diabetes detection, and subsequent sampling and data analysis methods. Experiments validated the accuracy of this system.

The remainder of this paper is organized as follows: Section II describes the breath analysis system in detail; Section III introduces the database and experiment config we used to test our system; Section IV is the experiment results and discussion and Section V concludes the paper.

II. SYSTEM DESCRIPTION

In this section, we will first introduce the structure of the device used to collect and measure human breath, followed by the key part of the device – the sensor array. After that we will describe how breath is collected and measured. Finally, a brief introduction will be given about the data analysis algorithms.

A. Structure of the Device

Fig. 1 is a snapshot of the internal structure of our device, whose main framework is shown in Fig. 2. The gas route consists of a vacuum pump and a gas room. Breath or fresh air is drawn from outside and pumped into the gas room. The gas room is a metal chamber with chemical sensors embedded in its shell. On contact with gas particles, the sensors' resistance will change. Each sensor's resistance is measured by a voltage divider circuit and transformed into raw voltage signals. The signals are then filtered by the signal processing circuit and collected by a data acquisition card. Finally they will be digitized and sent into a PC by a USB cable. We employed a fan for cooling to avoid accumulation of heat generated by the chemical sensors. A switching power supply is used to provide energy for the whole system. Table I lists some basic parameters of the proposed device.

The work is partially supported by the Natural Science Foundation of China (NSFC) (No. 61101150)

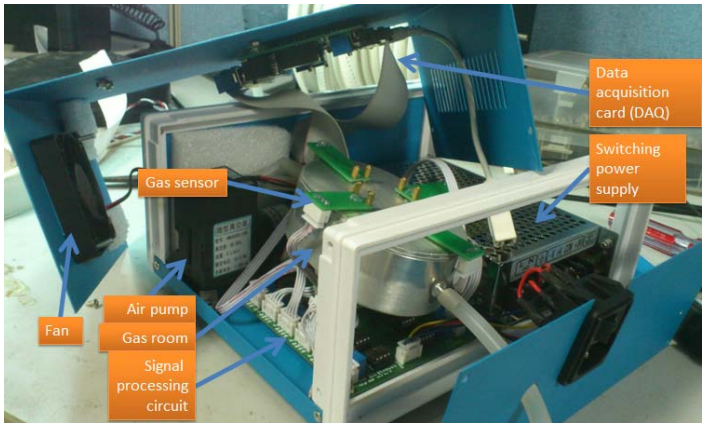

Figure 1. Internal structure of the proposed device.

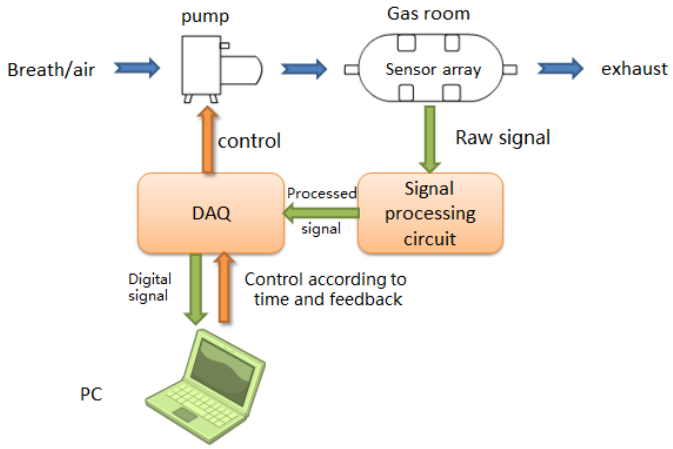

Figure 2. Main framework of the proposed device.

TABLE I. BASIC PARAMETERS OF THE PROPOSED DEVICE

Device parameters	Specifications
Size	22cm*18cm* 9cm
Gas room volume	130 mL
Sample frequency	8 Hz
Sampling duration	108 s
Injection speed	75 mL/s
Working temperature	25±10℃

B. *Sensor Array*

The sensor array is the key part of our device. The performance of the system is highly dependent on the capability of the sensors. We employed 7 commercial chemical sensors in the device, which are selected according to the experiment results in our previous work [8] and the results of simulating experiments with acetone. We also tried to use sensors with different measurement range and from different companies in order to form a complementary array. All 7 chemical sensors are metal oxide semiconductor (MOS) sensors. Most of them are sensitive to volatile organic compounds (VOCs) such as acetone. Reference [7] claimed that hydrogen sensing is also helpful for diabetes diagnosis, so we also included a sensor whose main function is detecting hydrogen. Furthermore, it is known that change in ambient temperature and humidity will lead to drift in MOS sensors' responses [9]. So a temperature-humidity sensor is also added to the array. Table II gives the detailed information of the sensors.

C. *Sampling Procedure*

The sampling of the proposed system is quite easy. The examinee is asked to exhale into a 600 mL Tedlar® gas bag through a disposable mouthpiece. Between the mouthpiece and the gas bag there is also an airtight box containing a small paper bag of silica gel, which is used to prevent part of the water vapor from getting into the gas bag. After breath collection, one can just plug the gas bag onto the connector of the device and click a button on the user interface of the PC software. Then the software will control the device to finish the measurement of the breath.

Fig. 3 shows sensor responses curves of a typical sample. Each sample consists of 4 stages, including:

1) Baseline stage (1 s): The baseline values of the sensors are recorded for future data preprocessing.

2) Injection stage (7 s): The pump opens, breath is drawn from the gas bag to the gas room at a constant speed.

3) Reaction stage (30 s): The pump is off, the sensors continue reacting with the gas particles.

4) Purge stage (70 s): The pump opens again, fresh air is drawn into the gas room to push the breath gas out. The software will stop recording the signals at the 70th second but the purging may continue until it finds that all the sensors have returned to their baseline.

D. *Data Analysis*

After it is obtained, a breath sample will first be preprocessed for further analysis. The aim of preprocessing is to compensate for drift and eliminate irrelevant information [7]. For each of the 7 chemical sensors, the baseline value is estimated by the average response of the baseline stage. Then the preprocessed response can be calculated by subtracting the baseline value from the corresponding sensor response.

TABLE II. DETAILED SENSOR INFORMATION

No.	Model	Function	Range	Company
1	TGS822	VOCs, CO	50-5000 ppm	Figaro Inc.
2	TGS2620	VOCs, H_2, CO	50-5000 ppm	
3	QS-01	VOCs, H_2, NH_3, CO	1-1000 ppm	FIS Inc.
4	TGS821	H_2	10-1000 ppm	Figaro Inc.
5	TGS2602	VOCs, H_2, NH_3, H_2S	1-30 ppm	
6	TGS826	NH_3,VOCs, H_2	30-300 ppm	
7	TGS2610-C00	VOCs, H_2	500-10000 ppm	
8	HTG3515 CH	Temperature	-40~110℃	Humirel Inc.
9		Humidity	10~95%RH	

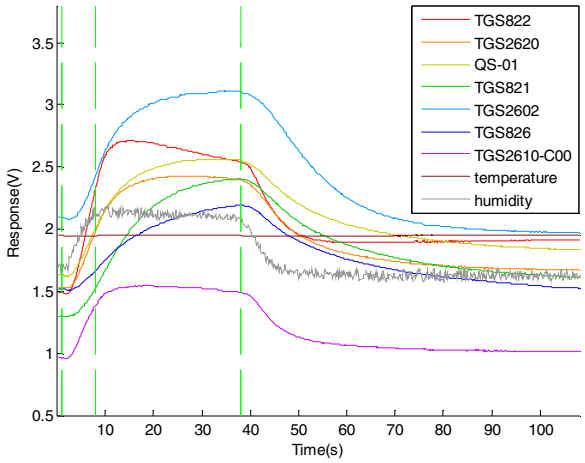

Figure 3. Typical sensor responses (not preprocessed) with 4 stages separated by green vertical lines.

Next, a sensor selection step will be carried out to select the optimal sensor array with the highest accuracy. Then we apply principle component analysis (PCA) on the responses of selected sensors to extract low-dimensional features. Finally, K-nearest neighbor (KNN) algorithm will be used to get the classification results. Here we set K = 3.

For the temperature sensor and the humidity sensor, we first use the conversion equations in the datasheet to transform their responses into temperature in centi degrees and relative humidity (%RH). Then, for temperature, we extract 1 feature for each sample, i.e. the average temperature during the whole sampling procedure. For humidity, we extract 2 features for each sample, i.e. the average humidity in the baseline stage, and the average humidity in the reaction stage. If these 2 sensors are used, their features will join the PCA feature extraction algorithm together with the responses of the chemical sensors.

III. EXPERIMENTS

A. Database Overview

In order to validate the efficacy of the proposed system, we collected breath samples of both healthy people and diabetics using the method introduced in Section II-C. Table III is a brief summary about the composition of the database. The healthy samples were collected in Guangdong Hospital of Traditional Chinese Medicine (Guangzhou, China, abbreviated as GDHTCM). The subjects' health states were confirmed by physical examinations. The diabetes samples were partly from inpatient volunteers in GDHTCM, and partly from outpatient volunteers in Hong Kong Institute of Diabetes and Obesity (The Chinese University of Hong Kong, Hong Kong SAR, China, abbreviated as HKIDO). The condition of the patients in HKIDO is better-controlled than the ones in GDHTCM.

TABLE III. COMPOSITION OF THE DATABASE

Subset	Location	Time	Type	Number
1	GDHTCM	2011.12~ 2012.1	Healthy	294
2	GDHTCM	2011.12~ 2012.1	Inpatient diabetes	117
3	HKIDO	2011.12	Outpatient diabetes	287

For each database subset, the average of the 7 chemical sensors' responses is shown in Fig. 4. They have been preprocessed by the method in Section II-D. We can see that the average response of the inpatient diabetes subset is much higher than that of the healthy subset, while the outpatient diabetes subset is only a little higher than the healthy subset. This can be observed more clearly in Fig. 5, which presents the average of max responses of the 7 chemical sensors for each subset. The error bars represent the standard deviation.

We have also recorded the body mass index (BMI) of some subjects in subset 3. BMI is defined as the individual's body mass divided by the square of his or her height. We consider a subject as obese if his or her BMI is larger than 27.5 kg/m^2. In subset 3, there are 106 nonobese diabetes samples and 173 obese ones. The rest samples' BMIs are unknown. Since breath acetone is a by-product of fat metabolizing, we will study if there is a difference of breath pattern between subjects with different BMI. However, there is no obvious distinction between the average responses of these two classes, so we will not show the curves here.

B. Sensor Selection

From Table II we know that all the 7 chemical sensors are sensitive to similar compounds, such as VOCs and hydrogen.

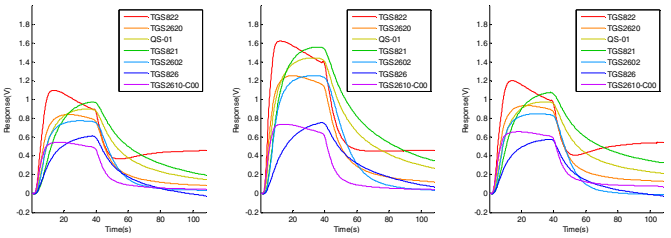

Figure 4. Average sensor responses (preprocessed) of 3 subsets (from left to right: subset 1,2,3).

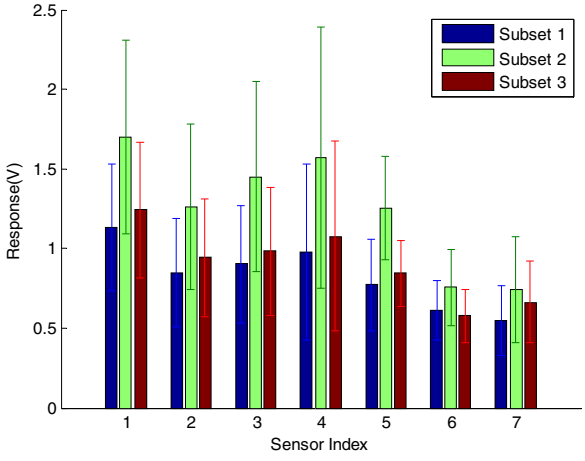

Figure 5. Average of max responses of the 7 chemical sensors for each subset. The error bars represent the standard deviation.

Although their sensitive spectrum may be a little different, there could be some sensors which provide largely redundant information and are useless for our task. Removing the useless sensors can not only improve the performance of the classifier, but also reduce the cost of the device [10]. It can also bring better understanding of the sensors and be inspirational for future system design.

Because our main task is diabetes diagnosis, we will use the accuracy of classification between healthy and diabetes samples as the evaluation criterion for our system. Now we have 7 chemical sensors, so there are 2^7-1=127 kinds of combinations. We will try all these combinations and find the top-ranking ones. This exhaustive method guarantees its result to be optimal. The time complexity is also acceptable. The accuracy of classification is computed using the strategy in Section III-C.

After selecting the optimal chemical sensor array, the features from the temperature and humidity sensor will be added to see if the accuracy could be further improved.

C. Classification Between Healthy and Diabetes Samples

In total there are 294 healthy samples and 404 diabetes samples in our database. The 2nd and 3rd subsets are merged. We will randomly take 147 healthy samples and 147 diabetes samples for training, and another 147+147 samples from both classes for testing. The accuracy is computed in 2 ways: sensitivity and specificity. Their definition can be found in [7]. This classification with random selected training and testing samples will be conducted 50 times and the average accuracy will be used as the final result.

D. Classification Between Nonobese and Obese Diabetes Samples

The strategy for nonobese and obese diabetes samples is similar. 53 nonobese diabetes samples and 53 obese ones are randomly taken for training. Another 53+53 samples from both classes are taken for testing. This procedure is done 50 times and average sensitivity and specificity are computed.

IV. Results and Discussion

A. Sensor Selection

The top 3 arrays for classifying between healthy and diabetes samples are shown in Table IV. They are ranked according to the average of sensitivity and specificity. In fact, the top 10 arrays all contain sensor 5 and 6, i.e. TGS2602 and TGS826. We can infer that these 2 sensors have the most importance. The conclusion is consistent with the results of our pilot simulating experiments, which showed these 2 sensors have the most sensitivity to acetone. Additionally, sensor 4 (TGS821) is not present in the top 10 arrays, indicating that it is of less importance for our task. When all 7 sensors are used, the accuracy is not so good compared to the top 1 array where only 3 sensors are used.

In order to assess the function of the temperature and humidity sensor, we added their features to the best array in Table IV. The average accuracy increases by 0.3% when humidity features are added but decreases by 0.3% when temperature feature is added. It means the humidity sensor is useful for our task, possibly because of the existence of water vapor in human breath.

TABLE IV. Comparison of Different Sensor Arrays

Rank	Sensors	Sensitivity	Specificity	Average
1	5,6,7	90.91%	89.78%	90.35%
2	1,2,5,6	88.76%	90.38%	89.57%
3	2,5,6,7	90.03%	88.41%	89.22%
21	1,2,3,4,5,6,7	87.29%	85.03%	86.16%

B. Classification Between Healthy and Diabetes Samples

According to the result of the last section, only sensor 5, 6, 7 and the humidity sensor are used for feature extraction. The classification algorithm is PCA coupled with KNN, as is introduced in Section II-D. Fig. 6 shows the distribution of the first 2 PCA coefficients of the 3 subsets. There is a clear separation between subset 1 and 2 but subset 3 has much overlap with subset 1. This possibly means that controlled diabetes samples are more similar to healthy samples than uncontrolled ones, as some researchers have found that the acetone concentration in breath of controlled diabetics is lower than uncontrolled ones, meanwhile close to healthy people [11]. The sensitivity, specificity and average accuracy are displayed in Table V.

C. Classification Between Nonobese and Obese Diabetes Samples

We try to tell if there is a difference between the breath sample of nonobese and obese diabetics. The distribution of the first 2 PCA coefficients of these 2 types of samples is shown in Fig. 7. In the figure, the samples lap over each other. The classification accuracy in Table V further confirmed that there is no clear distinction between these 2 types of samples. It should be noted that these samples are all from controlled diabetics. Further experiments need to be carried out to explore the relationship between BMI and breath pattern in diabetics not well controlled.

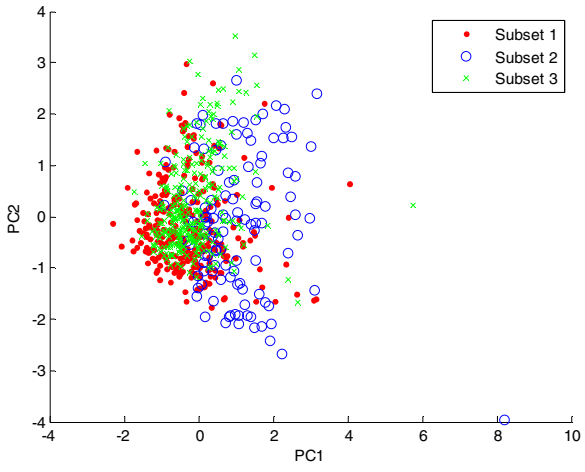

Figure 6. PCA 2-D plot for the 3 subsets of the database

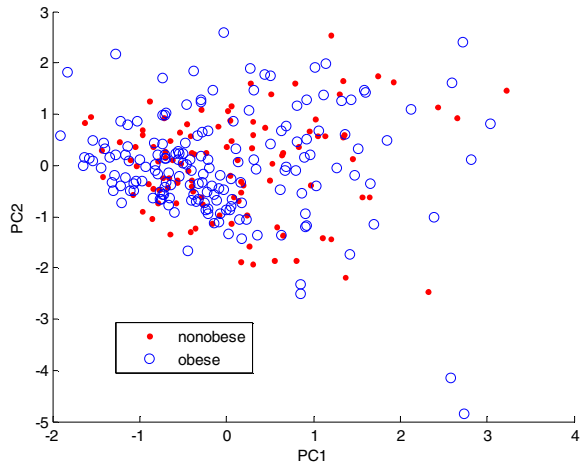

Figure 7. PCA 2-D plot for nonobese and obese diabetes samples

TABLE V. CLASSIFICATION RESULTS OF DIFFERENT SAMPLES

Type of subjects	Sensitivity	Specificity	Average
Healthy vs. Diabetes	91.43%	89.86%	90.65%
Nonobese vs. Obese	51.21%[a]	53.13%[b]	52.17%

a. The ratio for an obese sample to be classified correctly

b. The ratio for a nonobese sample to be classified correctly

V. CONCLUSION

This paper proposed a novel breath analysis system that is specialized for diabetes diagnosis. The device structure, sensor array, sampling procedure and data analysis method are introduced. An exhaustive sensor selection is done and several sensors are proved to have higher importance, including the humidity sensor. Experiments with breath samples from inpatient and outpatient diabetes are conducted to assess the accuracy of the system. The results show that the system is able to distinguish between healthy and diabetes samples with simple feature extraction and classification algorithms. We can also infer from the results that breath samples of controlled diabetics are more similar to samples of healthy people, while samples of uncontrolled diabetics have a clear difference from samples of healthy people. We also prove that there is no clear distinction between breath samples of nonobese and obese controlled diabetics. We hope this system could contribute a step toward a practical noninvasive diabetes screening system.

ACKNOWLEDGMENT

This work was supported by Guangdong Hospital of Traditional Chinese Medicine, Hong Kong Hospital Authority and Hong Kong Institute of Diabetes and Obesity under the Chinese University of Hong Kong.

REFERENCES

[1] C. Deng, J. Zhang, X. Yu, W. Zhang, and X. Zhang, "Determination of acetone in human breath by gas chromatography–mass spectrometry and solid-phase microextraction with on-fiber derivatization," Journal of Chromatography B, vol. 810, no. 2, pp. 269–275, 2004.

[2] C. Wang, A. Mbi, and M. Shepherd, "A study on breath acetone in diabetic patients using a cavity ringdown breath analyzer: Exploring correlations of breath acetone with blood glucose and glycohemoglobin a1c," IEEE Sensors Journal, vol. 10, no. 1, pp. 54–63, 2010.

[3] D. Smith, P. Spanel, A.A. Fryer, F. Hanna, and G.A.A. Ferns, "Can volatile compounds in exhaled breath be used to monitor control in diabetes mellitus?" J. Breath Res., vol. 5, no. 2, pp. 022001, 2011.

[4] A. Amann, G. Poupart, S. Telser, M. Ledochowski, A. Schmid, and S. Mechtcheriakov, "Applications of breath gas analysis in medicine," Int. J. Mass Spectrom., vol. 239, no. 2–3, pp. 227–233, 2004.

[5] Y. Lin, H. Guo, Y. Chang, M. Kao, H. Wang, and R. Hong, "Application of the electronic nose for uremia diagnosis," Sens. Actuators B: Chem., vol. 76, no. 1–3, pp. 177–180, 2001.

[6] J. Yu, H. Byun, M. So, and J. Huh, "Analysis of diabetic patient's breath with conducting polymer sensor array," Sens. Actuators B: Chem., vol. 108, no. 1–2, pp. 305–308, 2005.

[7] D. Guo, D. Zhang, N. Li, L. Zhang and J. Yang, "A novel breath analysis system based on electronic olfaction," IEEE Trans. on Biomedical Engineering, vol. 57, no. 11, pp. 2753-2763, 2010.

[8] D. Guo, D. Zhang, and L. Zhang, "An LDA based sensor selection approach used in breath analysis system," Sens. Actuators B: Chem., vol. 157, no. 1, pp. 265-274, 2011.

[9] F. Hossein-Babaei and V. Ghafarinia, "Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors," Sens. Actuators B: Chem., vol. 143, no. 2, pp. 641–648, 2010.

[10] J.W. Gardner, P. Boilot and E.L. Hines, "Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach," Sens. Actuators B: Chem., vol. 106, no. 1, pp. 114–121, 2005.

[11] P. Dalton, A. Gelperin and G. Preti, "Volatile metabolic monitoring of glycemic status in diabetes using electronic olfaction," Diabetes Technol Ther., vol. 6, no. 4, pp. 534-544, August 2004.