

Feature Extraction from Sensor Data for Detection of Wound Pathogen Based on Electronic Nose

Jia Yan*, Fengchun Tian, Qinghua He¹, Yue Shen¹,
Shan Xu, Jingwei Feng and Kadri Chaibou

College of Communication Engineering, Chongqing University, Chongqing, 400030, China
¹State Key Laboratory of Trauma, Burns and Combined Injury, Institute of Surgery Research,
Daping Hospital, Third Military Medical University, Chongqing, 400042, China

(Received November 2, 2010; accepted March 23, 2011)

Key words: electronic nose, wound infection, feature extraction, curve fitting

Three feature extraction methods, extraction from original response curves of sensors, curve fitting parameters and transform domains, for pathogen detection based on an electronic nose have been discussed. By using the integrals, coefficients of exponential fitting with two parameters and hyperbolic tangent function fitting, Fourier coefficients, and wavelet coefficients as features, 100% identification accuracy with the radical basis function neural network (RBFNN) classifier was reached for the seven pathogens. Theoretical analysis and experimental results indicate that the methods based on dynamic response are better than those based on steady response and can provide accurate identification of common pathogens present in wound infection.

1. Introduction

An electronic nose (enose), which is composed of an array of gas sensors as well as the corresponding pattern recognition algorithm, is able to imitate the olfactory system of humans and mammals, and is used for the recognition of gas and odor.⁽¹⁾ It plays a constantly growing role in general-purpose detectors of vapor chemicals in disease diagnosis, such as for lung diseases,^(2–4) diabetes,^(5,6) urinary tract infections,^(7,8) and bacteria detection,^(9–18) according to related articles published over the last few years.

Pathogen infection is one of the complications of wounds. Rapid and instant monitoring of the inflammatory response of wounds, especially the identification of the bacteria type and the phase of wound infection evolution, will help physicians to diagnose and choose the appropriate treatment. The application of the electronic nose

*Corresponding author: e-mail: yanjia119@163.com

discussed in this paper is to the detection of wound pathogens, and has the characteristics of being noninvasive, real-time, convenient, and highly efficient compared with traditional wound infection test methods.

The key point of wound pathogen recognition is the feature extraction method, by which the robust information from the sensor response curves is extracted with less redundancy, for the effectiveness of the subsequent pattern recognition algorithm.

Current feature extraction methods from sensor data can be approximately divided into three types. The first one extracts features from the original response curves of sensors, such as maximum values, integrals, differences, primary derivatives, secondary derivatives, the adsorption slope, and the maximum adsorption slope at a specific interval from the response curves. These features are extracted to discriminate or quantify six types of combustible gas and beverage,⁽¹⁹⁾ volatile organic compounds (VOCs),^(20,21) food,^(22–25) and bacteria.⁽²⁶⁾ The second one is based on curve fitting, which fits the response curves based on a specific model and extracts the set of fitting parameters as the features to classify bacteria⁽²⁷⁾ and 30 types of gas,⁽²⁸⁾ and to predict the concentration of hydrogen and ethanol and their mixture.⁽²⁹⁾ The third one is based on some transforms by extracting features in the transform domain, such as Fourier transform and wavelet transform, and then the transform coefficients are used as features to discriminate combustible gases^(30,31) and VOCs.^(32,33)

Previous works demonstrated that most applications of the electronic nose, including wound infection detection,^(16,17,34) only extract the steady-state response of sensors as the feature, without taking into account the other information in the entire response curves. Although the curve fitting method is adopted in some applications, the fitting models of the sensor response are not very diverse, and also, it is not used in wound infection monitoring. The methods of integrals, derivatives, Fourier transform, and wavelet transform show good performance in many applications,^(31–33) but they have not been applied well in wound infection monitoring. In general, there is no proper feature extraction method that has a good practical effect in wound infection monitoring, and how to extract robust features from experimental data for good recognition is a serious problem.

In this study, a gas sensor array of six metal oxide sensors and one electrochemical sensor is used to detect seven species of pathogen most common in wound infection: *P. aeruginosa*, *E. coli*, *Acinetobacter spp.*, *S. aureus*, *S. epidermidis*, *K. pneumoniae*, and *S. pyogenes*.^(18,35–40) We use an artificial neural network combined with various feature extraction methods to analyze information related to different pathogens on the basis of the gas sensor response curves. A fractional function model and hyperbolic tangent function model are used in curve fitting for feature extraction, and integrals, derivatives, Fourier transform, and wavelet transform are also used as the selected feature extraction methods in wound infection monitoring.

2. Methods and Materials

2.1 *Enose for wound pathogen detection*

In this study, the enose system for the detection of wound pathogens consists of wound headspace gas sampler, a sensor array, a signal conditioning circuit, and a data acquisition and processing unit. The array of sensors in the enose system, which is shown in Fig. 1, is composed of six metal oxide gas sensors, one electrochemical gas sensor, and temperature, humidity, and pressure sensors. The gas sensors are selected according to their sensitivity to the common volatile product in wound infection. More details on these gas sensors are shown in Table 1. To enhance the ability of the system to restrain environmental interference, a temperature sensor (LM35DZ), a humidity sensor (HIH4000), and a pressure sensor (SMI5552) are added into the sensor array to simultaneously collect the data of ambient parameters. The response signals of the headspace gas of the wound obtained by the sensor array are first conditioned through a conditioning circuit and then sampled and saved in a computer via a 14-bit data acquisition card (USB2002, Beijing art-control Inc.)

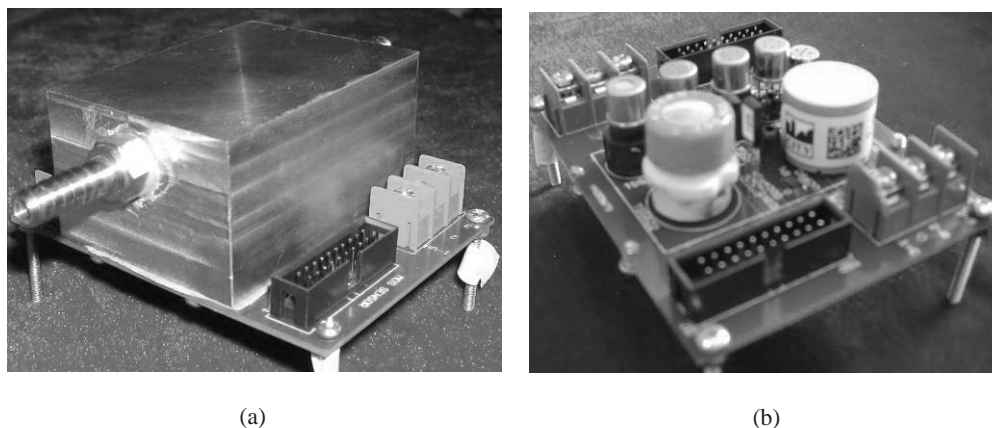


Fig. 1. Electronic nose (a) and its sensor array (b).

Table 1
Datasheet of gas sensor.

Sensor type	Typical analytes
TGS2600	Air contaminants, methane, carbon monoxide, isobutane, ethanol, hydrogen
TGS2602	VOCs and odorous gases, ammonia, H ₂ S, toluene, ethanol
TGS2620	Vapors of organic solvents, combustible gases, methane, CO, isobutane, hydrogen, ethanol
TGS822	Methane, carbon monoxide, isobutane, <i>n</i> -hexane, benzene, ethanol, acetone
QS-01	Hydrogen, CO, methane, isobutane, ethanol, ammonia
GSBT11	VOCs, HC, smoke, organic compounds
4ETO	Ethylene oxide, ethanol, methyl-ethyl-ketone, toluene, carbon monoxide

2.2 Sample preparation and measurement

Some studies^(18,35-40) revealed that the common pathogens responsible for wound infection include *P. aeruginosa*, *E. coli*, *Acinetobacter spp.*, *S. aureus*, *S. epidermidis*, *K. pneumoniae*, and *S. pyogenes*. These seven species of bacteria used in our tests were purchased from the Chinese National Institute for the Control of Pharmaceutical and Biological Products, and the National Center for Medical Culture Collection. The seven species of bacteria grow in media at 37°C with shaking at 150 rpm in a gyratory incubator shaker for 24 h. The culture medium is ordinary broth medium and the main components include peptone, NaCl, beef extract, and glucose. After 3 successive generations of subculture, the purchased bacteria became stable. Then they were inoculated into test agar slant. The headspace gas in each test slant that contains the metabolic products of bacteria is imported into the enose for the test through a Teflon tube by the gas flow rate control system and a pump.

The dynamic headspace method is adopted during all the experiments, and the process is as follows. The first stage is baseline collection, which is a 3 min pulse purge by injecting pure nitrogen into the sensor chamber. Then, the second stage (gas on) is adsorption collection, which involves a 3 min exposure of the sensor to the headspace air of bacteria conveyed into the sensor chamber by a pump. The third stage (gas off) is the desorption stage, which takes 4 min, where pure nitrogen is conveyed into the sensor chamber. At the end of each stage, prior to the next stage, 10 min purging of the sensor chamber using high-purity nitrogen is performed. The gas flow is controlled by a gas flow rate control system, which contains a rotor flowmeter, a pressure-retaining valve, a steady flow valve, and a needle valve. The flow rate was kept at 50 ml/min. Ten repetitions for each pathogen, that is, 70 experiments for all seven pathogens under the same conditions, are carried out, and so 70 samples are collected.

3. Feature Extraction

3.1 Signal preprocessing

Several methods have been adopted to preprocess the sensor signals, such as the ones listed in Table 2.

The difference method can usually eliminate the additive errors, which are added both to the baseline V_{ij}^{\min} and the steady-state response V_{ij}^{\max} (V_{ij} means the response of

Table 2
Preprocessing methods.

Method	Formula
Difference	$x_{ij} = V_{ij}^{\max} - V_{ij}^{\min}$
Relative difference	$x_{ij} = V_{ij}^{\max} / V_{ij}^{\min}$
Fractional difference	$x_{ij} = (V_{ij}^{\max} - V_{ij}^{\min}) / V_{ij}^{\min}$
Log difference	$\lg(V_{ij}^{\max} / V_{ij}^{\min})$
Normalization	$x_{ij} = (V_{ij} - V_{ij}^{\min}) / (V_{ij}^{\max} - V_{ij}^{\min})$

the i th sensor to the j th gas). To reduce multiplicative error, we usually use the relative difference methods that calculate the ratio of the steady-state response to the baseline. The relative difference and fractional difference methods are helpful to compensate the temperature influence on the sensors, and the fractional difference method can linearize the relationship between the resistance of a metal-oxide sensor and odor concentrations. This method provides good recognition results for the discrimination of several types of coffee using neural networks.⁽⁴¹⁾

The log difference method is suitable when the variation of the concentration of the odor-producing material is very large because it is able to linearize the highly nonlinear relationship between odor concentration and sensor output. The last method is normalization, which limits each sensor output to between 0 and 1, thereby keeping each element of the response vector in the same order. Not only can it reduce the calculation error of stoichiometric recognition, but it is also very effective when not odor concentration but the precise identification of the odor is of interest.

3.2 Feature extraction

The purpose of feature extraction is to extract features from multidimensional sensor signals to obtain an optimal recognition result. The features include the steady-state signal, the transient signals, the parameters of different response curve fitting models, and the coefficients of various transforms. A brief description of the parameters is given in Table 3.

The features are divided into three types. The first type of feature is extracted from the original response curves of sensor output such as the maximum response, integrals, and derivatives in Table 3. The second type consists of parameters estimated from different models of the transient response, such as polynomial functions, exponential functions, fractional functions, arc tangent functions, and hyperbolic tangent functions, which fit the response curves. All the curve fittings are performed using least squares minimization. The third type is based on some transforms of the output, and the transform coefficients are extracted as the features, such as the coefficients of Fourier transform and wavelet transform.

The integrals and derivatives of signals have specific physical meanings. They both reflect different information of the reaction kinetics at different aspects. Integrals may represent the cumulative total of the reaction degree change and derivatives represent the rate of the reaction.⁽¹⁹⁾

Instead of extracting characteristics from the raw data curve, an alternative method is to model sensor response curves. Discriminations are then performed based on the model coefficients. Curve fitting is a data processing method that approximately characterizes or analogizes the functional relationship of discrete points in the plane using continuous curves. The curve fitting method approximates discrete data using analytical expressions. The basic idea is to determine the fitting function in accordance with the trend of the discrete points. It utilizes the discrete points, but is not limited to these discrete points. In curve fitting, the critical and complex problem is how to select and envision the specific form of the unknown function. There are no strict mathematical laws or theories to follow for curve fitting. However, two approaches are generally

Table 3
Brief description of the parameters extracted from the sensor response curves.

Parameter	Description
Maximum response	Max (sensor value)
Integral	$I = \int_a^b f(x)dx$, $I = \int_a^b f(x)$: sensor value, x : time from a to b , a : the beginning of the adsorption stage, b : the end of the desorption stage
Derivative	Difference between two samples during measurement
Polynomial function	$Y = A_0 + A_1x + A_2x^2 + A_3x^3$ On: $Y = (\text{sensor value-baseline})$, $x = \text{time from gas On to gas Off}$ Off: $Y = (\text{final response-sensor value})$, $x = \text{time from gas Off to the end}$
Fractional function	$Y = x/Ax+B$, where Y and x are defined as in the polynomial fit
Exponential function 1	$Y = A(1-\exp(-x/T))$, where Y and x are defined as in the polynomial fit
Exponential function 2	$Y = A_0 + A_1(1-\exp(-x/T_1)) + A_2(1-\exp(-x/T_2))$ where Y and x are defined as in the polynomial fit
Arctangent function	$Y = A \times \arctan(x/B)$, where Y and x are defined as in the polynomial fit
Hyperbolic tangent function	$Y = A \times \tanh(x/B)$, where Y and x are defined as in the polynomial fit
Fourier descriptors	Coefficients of the FFT of sensor value
Wavelet descriptors	Coefficients of the DWT of sensor value

followed: (1) determine the basic type of function through the study of the physical concepts between the variables and deep understanding of professional knowledge and (2) determine the type of function through the observation of general trends of the curve of experimental data. Several coefficients that fit the response curves are estimated using different models of the transient response, such as the polynomial functions, exponential functions, fractional functions, arc tangent functions, and hyperbolic tangent functions, and are used as parameters to characterize each measurement. The parameter extraction and curve fitting are carried out on both the ascending and descending parts of the response curves. All the curve fittings are performed using least squares minimization.

The widely used Fourier transform, for which the basis functions are sines and cosine, maps the original data into a new space. Owing to the fact that the basis functions of Fourier transform are defined in an infinite space and are periodic, Fourier transform is best suited to signals with the same features.⁽³¹⁾ It decomposes the original response into a superposition of the DC component and different harmonic components. Then, using the amplitudes of these components as features, qualitative as well as quantitative analysis can be realized. Wavelet transform is an extension of Fourier

transform. It maps the signals into a new space with basis functions quite localizable in time and frequency space. They are usually of compact support, orthogonal/biorthogonal and have proper Lipschitz regularity and vanishing moment. The wavelet transform decomposes the original response into the approximation (low frequencies) and details (high frequencies). It bears a good anti-interference ability for the following pattern recognition to use the wavelet coefficients of certain sub-bands as features.

For simplicity, we select the Daubechies (db) series compact support orthogonal nonsymmetric wavelet for feature extraction. Gas sensor signals are affected by various types of noise, which appear mostly in high-frequency components.^(42,43) When the decomposition level is too low, sub-band coefficients will retain more high-frequency components, which is not conducive to suppressing the noise. After a test comparison, the db6 wavelet is selected for wound pathogen detection,⁽⁴⁴⁾ and the approximating coefficients, which are used as the initial feature, are obtained after a wavelet decomposition of 12 levels is carried out for the original signal.

4. Results and Discussion

With the proposed methods, we compare the pattern recognition results of these features, extracted from the original response curves of sensors, parameters of curve fitting, and transform domains. Analysis has been carried out using the RBFNN classifier. Five samples of each pathogen are taken as training samples, and the other five samples of each type of pathogen are taken as testing samples. Thus, the training set and testing set each have 35 samples.

4.1 Effects of different preprocesses on steady-state signal of sensors

RBFNN identification results when considering the maximum response of sensors are shown in Table 4.

None of the five preprocessing methods can provide a 100% correct classification rate in the discrimination of the commonly encountered pathogens in wound infection. In species classification, 94.29% of the samples (33/35) in the normalization method are correctly classified, which is the highest achieved, while only 77.14% of the samples (27/35) in the difference method are correctly discriminated, which it is the lowest achieved. The correct classification rates of the relative difference, fractional difference, and log difference methods are between these two values. The results of RBFNN imply

Table 4
Identification results of five preprocessing methods.

Preprocessing method	Recognition result	Recognition rate (%)
Difference	27/35	77.14
Relative difference	28/35	80
Fractional difference	28/35	80
Log difference	30/35	85.71
Normalization	33/35	94.29

that among all the feature extraction methods, only the method of extracting steady-state response as a feature for recognizing the wound pathogen cannot achieve a good result. The steady-state response of the sensor is only one point on the response curve; it only contains static information and does not contain enough information to correctly classify the samples, such as the dynamic information included in the whole dynamic response curve. It was also found that the normalization method provides a better identification result than the other four methods; only two samples were not correctly classified by it. This means that the normalization method is more suitable for preprocessing the sensor data for wound pathogen detection if only features from the steady-state response are considered, because it can eliminate the effect of a concentration difference on recognition.

4.2 Integral and derivative methods

The integrals and differences reflect different information of the sensor reaction process at different aspects when it is exposed to odors. Because the steady-state response of sensors cannot provide good classification of pathogens, dynamic response has also been investigated as a feature extraction method.

Integrals may represent the accumulative total of the reaction degree change and derivatives may represent the rate of the sensor reaction to the odor. The integral and derivative expressions are given in eqs. (1) and (2).

$$\text{Integrals: } I = \int_a^b f(x) dx \quad (1)$$

Here, $f(x)$ is the sensor response value, a represents the start time of the headspace air of bacteria flowing into the sensor chamber, b is the end of the recovery time, and x represents a time point between a and b .

$$\text{Derivative: } f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \quad (2)$$

Here, x_i is the time point from a to the start time of the desorption stage, and $f(x_i)$ and $f(x_{i+1})$ represent the sensor values at x_i and x_{i+1} time points, respectively.

For integrals, the extracted feature is just the area between the sensor response curve and the time axis from a to b . The Newton-Cotes method is applied to compute the area with a simpler approximate function. The Cotes formula yields the area under a parabola that passes through five points that are equally spaced (i.e. x_0, x_1, x_2, x_3, x_4) on a curve. Thus, the area is computed as follows:

$$I = \int_a^b f(x) dx \approx \frac{b-a}{90} [7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)], \quad (3)$$

$$x_k = a + k \frac{b-a}{4}, \quad (k = 0, 1, 2, 3, 4).$$

For derivatives, we compute the mean derivative over intervals of 180 samples; thus, for the whole transient response, 10 features are obtained.

The identification results of the integral and derivative methods classified using an RBFNN are given in Tables 5 and 6, respectively. Here, the identification result can be expressed as

$$Result = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \cdots & x_{ij} & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mm} \end{bmatrix}, \quad (4)$$

where x_{ij} denotes the number of samples in class i judged to belong to class j (the judgment is correct when $i = j$, and incorrect otherwise), and m is the total number of classes in the test set.

In Table 5, it is interesting to note that the result of the integral method is much better than that of the traditional steady-state response method. In the data analysis, a 100% correct classification rate is achieved in the discrimination of the seven species of

Table 5
Discrimination result of interval.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1, *E. coli*; 2, *K. pneumoniae*; 3, *P. aeruginosa*; 4, *Acinetobacter*; 5, *S. aureus*; 6, *S. epidermidi*; 7, *S. pyogene*.

Table 6
Discrimination result of derivative.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	1	0	4	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

pathogen, which means that this method may produce much more informative features compared with the results obtained with the steady-state methods. In Table 6, only one sample of category 4 is classified into category 2 when using the derivative method; thus, a 97.14% correct classification rate is achieved in the discrimination of the seven species. Compared with all the steady-state methods, the derivative method has shown much better results. This means that the dynamic response methods, such as the integral and derivative methods, provide more useful information than the steady-state response methods; that is, they have better results in pattern recognition.

4.3 Curve fitting parameters

The RMSE of curve fitting using the six aforementioned algorithms are shown in Table 7. In Table 7, the three-order polynomial model has the least errors, resulting in the best overall fitting. Exponential function 1 gives worse fitting than exponential function 2. This is explained by the larger number of coefficients in the exponential 2 fitting. The fractional function, the arctangent function and the hyperbolic tangent do not have very different fit errors, and they have much worse fits than the three-order polynomial function and exponential 2.

The quality of curve fitting depends on the type of model and number of parameters in the model. Increasing the number of parameters in curve fitting generally gives a better fitting. The type of basic function also affects the quality of curve fitting. Discrimination results of curve fitting coefficients, using the RBFNN classifier, are given in Tables 8 to 13.

From Tables 8 to 13, we can see that the recognition results of using curve fitting

Table 7

Root-mean-square error calculated for the six different response curve fitting algorithms.

Fit method	Three-order polynomial	Fractional function	Exponential 1	Exponential 2	Arctangent	Hyperbolic tangent
RMSE	0.5643	1.5650	1.7317	0.8617	1.6396	1.7822

Note: RMSE values were calculated using data from all 70 experiments.

Table 8

Discrimination result of the three- order function model.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	4	0	1	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	1	0	4	0
7	0	0	0	1	0	0	4

1–7 are the same as in Table 5.

Table 9
Discrimination result of fractional polynomial function model.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	1	4	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

Table 10
Discrimination result of exponential function model 1.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

Table 11
Discrimination result of exponential function model 2.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	2	1	1	0	1	0
3	0	2	2	1	0	0	0
4	0	1	0	4	0	0	0
5	0	0	0	0	5	0	0
6	0	0	3	0	0	2	0
7	0	0	0	1	0	0	4

1–7 are the same as in Table 5.

coefficients as the feature are generally better than those of steady-state response except for the case of exponential function fitting with five parameters to be estimated. Because the fitting coefficients contain not only the maximum response curves but also all the information in the whole phase of the response curves, in contrast to the steady-state

Table 12
Discrimination result of arc tangent function model.

	1	2	3	4	5	6	7
1	4	0	0	1	0	0	0
2	0	4	1	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	1	0	0	0	0	4

1–7 are the same as in Table 5.

Table 13
Discrimination result of hyperbolic tangent function model.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

response method, they can reflect the sensor reaction process better. The information on the sensor reaction process is of great importance in pattern recognition.

The exponential function with 2 parameters and the hyperbolic tangent function fitting coefficients both have the best recognition effect and achieve a 100% correct classification rate. Although models with more parameters, such as the exponential function with five parameters and three-order polynomial function, have better fitting results and the corresponding fitting errors are smaller, the results of final identification are not better than those of models with fewer parameters, and the identification rates are only 68.57 and 91.43%, respectively. Although the fractional function model and arc tangent function model have 97.14 and 91.43% correct classification rates, respectively, they have fewer parameters to estimate. Therefore, we cannot determine the identification result as being good or not just from the fitting effect or from the number of parameters. Also, one cannot confirm that a small fitting error can guarantee better recognition results. Increasing the number of parameters in curve fitting generally yields better fitting to a certain extent, but too many parameters will make the model too complex, and using such a model would incur the risk of overfitting the data, resulting in a poor generalization capability of the model and decrease in identification ability.

4.4 Fourier transform and wavelet transform

Fourier transform decomposes the original response curve into a superposition of the DC component and different harmonic components; then, the amplitudes of the DC component as optimal features are sent into the RBFNN classifier. The RBFNN discrimination results of Fourier transform descriptors are given in Table 14. Wavelet transform decomposes the original signal into approximation and detail parts, and wavelet coefficients of a specific sub-band are extracted as features. Wavelet transform has high anti-interference ability. The approximating coefficients, which are used as the optimal feature, are obtained after a wavelet decomposition of 12 levels is carried out for the original signal. Then, the mean of the selected coefficients is sent into the RBFNN classifier. The RBFNN discrimination results of wavelet transform descriptors are given in Table 15.

Obviously, Fourier transform and wavelet transform are more suitable than traditional means of feature extraction for the discrimination of commonly encountered pathogens in wound infection. The two sets of features are shown to significantly enhance the performance of subsequent classification algorithms. It is noteworthy that either Fourier descriptors or wavelet descriptors can achieve a 100% discrimination rate.

Figure 2 presents box plots, which show the selected features that achieve 100% correct classification rate and the maximum features of the seven classes of the pathogen samples. The box plot for each class contains information on the mean, quartile value, and outliers of all the samples in this class, revealing the interclass distance of a class. The distribution across the 7 box plots indicates the interclass distances between different classes. The more dispersed the box plots are, the greater the differences between classes. From Figs. 2(a) to 2(f), it is evident that for samples in the same class, the degree of scatter of the selected feature is smaller than that of the maximum feature. For samples in different classes, the degree of scatter of the selected feature is much greater than that of the maximum feature, and this means that the discrimination ability of the selected features is much stronger than that of the maximum feature.

The recognition effects of all the feature extraction methods in identifying commonly encountered pathogens in wound infection are shown in Table 16.

Table 14
Discrimination result of Fourier descriptors.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

Table 15
Discrimination result of wavelet descriptors.

	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	5	0	0	0	0	0
3	0	0	5	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	5	0	0
6	0	0	0	0	0	5	0
7	0	0	0	0	0	0	5

1–7 are the same as in Table 5.

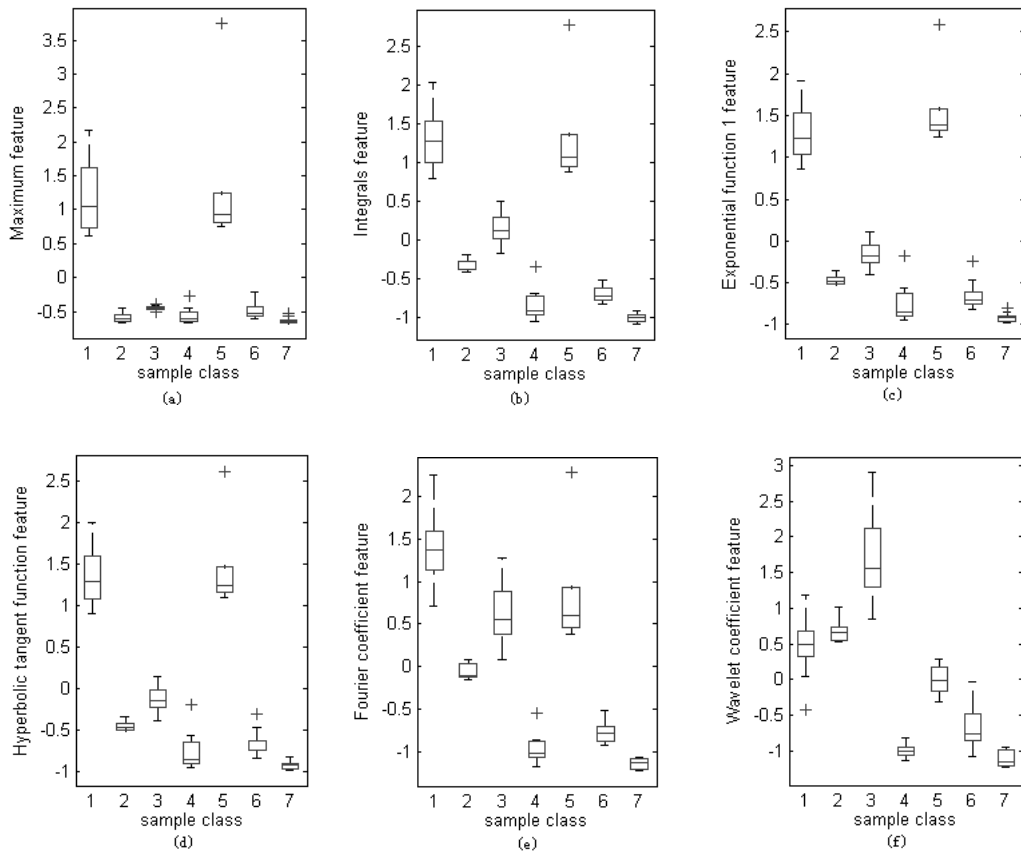


Fig. 2. Box plots for the maximum feature (a) and selected features (b)–(f).

Table 16
Results from neural network recognition.

Preprocessing method	Recognition rate (%)
Difference	77.014
Relative difference	80
Fractional difference	80
Log difference	85.71
Normalization	94.29
Integrals	100
Derivatives	97.14
Three-order polynomial function	91.43
Fractional function	97.14
Exponential function 1	100
Exponential function 2	68.57
Arctangent function	91.43
Hyperbolic tangent function	100
Fourier coefficient	100
Wavelet coefficient	100

5. Conclusions

In this work, we discussed the effect of various feature extraction methods of sensor signals on the final result of pathogen recognition in terms of the application of an electronic nose in detecting bacteria in human wound infection. The various features selected were then used as inputs to a nonlinear RBF artificial neural network to classify the pathogens. The traditional e-nose signal feature extraction methods often use the steady-state response as a feature, which seems to yield good results in other applications of e-nose, but is not suitable in the detection of pathogens found in wound infection. Pathogens of wound infection have complex metabolites, and their concentrations are low, so it is difficult to obtain good results only by adopting one feature of steady-state response. Gas sensor response is a complicated dynamic process, and the whole response curve contains much useful information for pattern recognition. Therefore, using transient information can improve the recognition ability. Three ways of extracting features were used: the original response curve, curve fitting parameters, and the transform domain. In the original response curve, processing the steady-state response using the sensor normalization method is more effective than using the other preprocessing methods. The recognition rates of integral and derivative methods increase greatly in comparison with the steady-state response feature and the integral methods have a 100% classification rate. The curve fitting method is also more effective than traditional methods, and the models with more fitting parameters have worse results than the models with fewer fitting parameters. The exponential function model with two parameters and the hyperbolic tangent function model both have the best classification

rate of 100%. For the case of the transform domain, the selected Fourier transform and wavelet transform can also achieve a 100% correct recognition effect.

Acknowledgements

This work is supported by the Natural Science Foundation Project of CQ CSTC, 2009, P.R. China, China Postdoctoral Science Foundation funded project (20090461445), Chongqing University Postgraduates' Science and Innovation Fund, Project Number: 200911B1A0100326, and Chongqing University Postgraduates' Innovation Team Building Project, Team Number: 200909 C1016.

References

- 1 E. L. Hines, E. Llobet and J. W. Gardner: *IEE Proc. Circ. Dev. Syst.* **146** (1999) 297.
- 2 D. T. V. Anh, W. Olthuis and P. Bergveld: *Sens. Actuators, B* **111** (2005) 494.
- 3 C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D'Arcangelo, C. Roscioni, A. Finazzi-Agro and A. D'Amico: *Biosens. Bioelectron.* **18** (2003) 1209.
- 4 M. Phillips, R. N. Cataneo, A. R. C. Cummin, A. J. Gagliardi, K. Gleeson, J. Greenberg, R. A. Maxfield and W. N. Rom: *Chest* **6** (2003) 123.
- 5 N. Makisimovich, V. Vorotyntsev, N. Nikitina, O. Kaskevich, P. Karabun and F. Martynenko: *Sens. Actuators, B* **35** (1996) 419.
- 6 J.-B. Yu, H.-G. Byun, M.-S. So and J.-S. Huh: *Sens. Actuators, B* **108** (2005) 305.
- 7 Y.-J. Lin, H.-R. Guo, Y.-H. Chang, M.-T. Kao, H.-H. Wang and R.-I. Hon: *Sens. Actuators, B* **76** (2001) 177.
- 8 A. K. Pavlou, N. Magan, C. McNulty, J. M. Jones, D. Sharp, J. Brown and A. P. F. Turner: *Biosens. Bioelectron.* **17** (2002) 893.
- 9 V. Rossi, R. Talon and J.-L. Berdague: *J. Microbiol. Methods* **24** (1995) 183.
- 10 T. D. Gibson, O. Prosser, J. N. Hulbert, R. W. Marshall, P. Corcoran, P. Lowery, E. A. Ruck-Keene and S. Heron: *Sens. Actuators, B* **44** (1997) 413.
- 11 J. W. Gardner, M. Craven, C. Dow and E. L. Hines: *Meas. Sci. Technol* **9** (1998) 120.
- 12 J. W. Gardner, H. W. Shin and E. L. Hines: *Sens. Actuators, B* **70** (2000) 19.
- 13 P. Lykos, P. H. Patel, C. Morong and A. Joseph: *J. Microbiol.* **9** (2001) 213.
- 14 R. Dutta, E. L. Hines, J. W. Gardner and P. Boilot: *Biomed. Eng. Online* **16** (2002) 1.
- 15 R. Dutta and R. Dutta: *Sens. Actuators, B* **120** (2006) 156.
- 16 A. Setkus, A.-J. Galdikas, Z.-A. Kancleris and A. Olekas: *Sens. Actuators, B* **115** (2006) 412.
- 17 A. L. P. S. Bailey, A. M. Pisanelli and K. C. Persaud: *Sens. Actuators, B* **131** (2008) 5.
- 18 A. Setkus, Z. Kancleris, A. Olekas, R. Rimdeik, D. Senuliene and V. Strazdiene: *Sens. Actuators, B* **130** (2008) 351.
- 19 S. Zhang, C. Xie, D. Zeng, Q. Zhang, H. Li and Z. Bi: *Sens. Actuators, B* **124** (2007) 437.
- 20 E. Llobet, J. Brezmes, X. Vilanova, J. E. Sueiras and X. Correig: *Sens. Actuators, B* **41** (1997) 13.
- 21 A. Setkus, A. Olekas, D. Senuliene, M. Falasconi, M. Pardo and G. Sberveglieri: *Sens. Actuators, B* **146** (2010) 539.
- 22 S. Roussel, G. Forsberg, V. Steinmetz, P. Grenier and V. Bellon-Maurel: *J. Food Eng.* **37** (1998) 207.
- 23 S. Balasubramanian, S. Panigrahi, C. M. Logue, H. Gu and M. Marchello: *J. Food Eng.* **91** (2009) 91.

- 24 I. Concina, M. Falasconi, E. Gobbi, F. Bianchi, M. Musci, M. Mattarozzi, M. Pardo, A. Mangia, M. Careri and G. Sberveglieri: *Food Control* **20** (2009) 873.
- 25 N. E. Barbri, J. Mirhisse, R. Ionescu, N. E. Bari, X. Correig, B. Bouchikhi and E. Llobet: *Sens. Actuators, B* **141** (2009) 538.
- 26 G. C. Green, A. D. C. Chan, H. Dan and M. Lin: *Sens. Actuators, B* **152** (2011) 21
- 27 M. Holmberg, F. Gustafsson, E. G. Hornsten, F. Winqvist, L. E. Nilsson, L. Ljung and I. Lundstrom: *Biotechnol. Tech.* **12** (1998) 319.
- 28 L. Carmel, S. Levy, D. Lancet and D. Harel: *Sens. Actuators, B* **93** (2003) 67.
- 29 T. Eklov, P. Mirtensson and I. Lundstrijm: *Anal. Chim. Acta* **353** (1997) 291.
- 30 H. Ge, H. Ding and J. Liu: The 13th International Conference on Solid-State Sensors, Actuators and Microsystems (Seoul, Korea, 2005).
- 31 R. Ionescu and E. Llobet: *Sens. Actuators, B* **81** (2002) 289
- 32 C. Distanto, M. Leo, P. Siciliano and K. C. Persaud: *Sens. Actuators, B* **87** (2002) 274.
- 33 E. Llobet, J. Brezmes, R. Ionescu, X. Vilanova, S. Al-Khalifa, J. W. Gardner, N. Barsan and X. Correig: *Sens. Actuators, B* **83** (2002) 238
- 34 K. C. Persaud, A. M. Pisanelli, A. Bailey, M. Falasconi, M. Pardo, G. Sberveglieri, D. Senulienė, A. Šetkus, R. Rimdeika, K. Dunn, M. Gobbi and U. Schreiter: *IEEE SENSORS 2008 Conference*.
- 35 J. N. Labows, K. J. Mcginley, G. Webster and J. J. Leyden: *J. Clin. Microbiol.* **12** (1980) 521.
- 36 A. Giacometti, O. Cirioni, A. M. Schimizzi, M. S. Del Prete, F. Barchiesi, M. M. Derrico, E. Petrelli and G. Scalise: *J. Clin. Microbiol.* **2** (2000) 918.
- 37 R. A. Allardyce, A. L. Hill and D. R. Murdoch: *Diagn. Micr. Infec. Dis.* **55** (2006) 255.
- 38 T. D. Gibson, O. Prosser, J. N. Hulbert, R. W. Marshall, P. Corcoran, P. Lowery, E. A. Ruck-Keene and S. Heron: *Sens. Actuators, B* **44** (1997) 413.
- 39 C. M. MC Entegart: *Detection and discrimination of enteric bacteria using gas sensor array*, PhD thesis, May, 2003.
- 40 H. Trill: *Diagnostic technologies for wound monitoring*, PhD thesis, March, 2006.
- 41 J. W. Gardner, H. V. Shumer and T. T. Tan: *Sens. Actuators, B* **6** (1992) 71.
- 42 F. Tian, S. X. Yang and K. Dong: *Sensors* **5** (2005) 85
- 43 F. Tian, S. X. Yang and K. Dong: 2005 IEEE International Conference on Networking, Sensing and Control (Tucson, Arizona, U.S.A, 2005).
- 44 T. Fengchun, X. Xuntao, Y. Shen, Y. Jia, M. Jianwei and L. Tao: *Sens. Mater.* **21** (2009) 155.