Available online at www.sciencedirect.com
SCIENCE DIRECT®
Sensors and Actuators B 93 (2003) 67–76
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

A feature extraction method for chemical sensors in electronic noses

L. Carmel[a,*], S. Levy[b], D. Lancet[c], D. Harel[a]

[a]*Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel*
[b]*Citala Ltd., Ramat Hachayal Tel-Aviv 69710, Israel*
[c]*Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel*

Abstract

We propose a new feature extraction method for use with chemical sensors. It is based on fitting a parametric analytic model of the sensor's response over time to the measured signal, and taking the set of best-fitting parameters as the features. The process of finding the features is fast and robust, and the resulting set of features is shown to significantly enhance the performance of subsequent classification algorithms. Moreover, the model that we have developed fits equally well to sensors of different technologies and embeddings, suggesting its applicability to a diverse repertoire of sensors and analytic devices.

Keywords: Feature extraction; Electronic nose; Curve fitting; Quartz-microbalance sensors; Metal-oxide sensors

1. Introduction

Electronic noses (or, in short, eNoses) are analytic devices that play a constantly growing role as general purpose detectors of vapor chemicals [1]. The main component of an eNose is an array of non-specific sensors, i.e., sensors that interact with a broad range of chemicals with varying strengths. Correspondingly, an analyte stimulates many of the sensors in the array and elicits a characteristic response pattern. The sensors inside an eNose are made of diverse technologies. Depending on the type of sensor, a certain physical property is changed as a result of an exposure to gaseous analytes. During the measurement process a signal is obtained by constantly recording the value of this physical property.

The signals should then be analyzed for the benefit of the specific application. The vast majority of the applications involve a *classification* process—identifying an unknown sample by comparing its pattern with those of known analytes. For details on this kind of pattern recognition, see, e.g., [2]. To list a few examples, eNoses are used for quality assessment of food products [3–5], for medical diagnostics [6,7], and for determining the amount of off-odor in packaging materials [8].

Whatever classification algorithm is used, it requires the measured signals as input. However, since a typical signal is comprised of a few hundred measured values, a preceding stage of *feature extraction* is frequently required. This is the process of finding a small set of parameters that somehow represent the entire signal. To date, a small group of feature extraction methods is used by the vast majority of the community, all capture only a portion of the information contained in the signals. Even though these methods are satisfactory for some applications, it is generally accepted that performance can be enhanced by the use of more optimal methods. Yet, no practical alternatives to the currently used methods have been proposed.

In this paper we present a new feature extraction method that extracts much more information from the signals, yet keeps the number of features small. The idea that underlies the method is to model the time-dependency of the response by an analytic expression, which is completely characterized by a small set of parameters; these parameters are then taken to be the features. For every measurement we find the corresponding values of the features by carrying out a curve-fitting procedure.

The analytic model, which we henceforth call the *Lorentzian model*, is derived from a very simple physical description of the measurement process. It uses four parameters, all with a precise physical meaning, that are obtained from a fast and robust curve-fitting process. We show that using them as features in later classification tasks results in a

* Corresponding author.
E-mail addresses: liran.carmel@weizmann.ac.il (L. Carmel), slevy67@yahoo.com (S. Levy), doron.lancet@weizmann.ac.il (D. Lancet), dharel@weizmann.ac.il (D. Harel).

0925-4005/03/$ – see front matter
doi:10.1016/S0925-4005(03)00247-8

significantly improved classification rate, suggesting their utility as the input for data analysis algorithms.

The shape of a signal depends on the type of sensor, the type of stimulus, the physical arrangement of the apparatus, and the way by which the stimulus is introduced to the sensor. Any of these can be dramatically varied between experiments, resulting in a diverse repertoire of signals. Surprisingly enough, our model shows excellent robustness with respect to changing these parameters. Specifically, we have successfully tested the model against two different sensor modules differing by the type of sensors (quartz-microbalance (QMB) sensors and metal-oxide (MOX) sensors), by their physical arrangement, and by the shape and volume of their housing chamber. Relying upon these results, we speculate that our model has a broad range of applicability, being valid for many as of yet unexamined sensor technologies and chamber designs. Probably, it would also be valid for other kinds of analytic devices that contain chemical sensors.

Besides the Lorentzian model, we introduce two alternative models that are slightly inferior in general, but may be found more beneficial in certain applications. The first, which we call the *exponential model*, is derived by slightly changing the physical assumptions that led to the Lorentzian model. It also uses four physically interpretable and rapidly computable parameters, but its fit to the measured signals is slightly worse than that of the Lorentzian model. The second, the *double-sigmoid model*, is purely empirical. It has the best fit with the measured signals, but at the expense of using nine parameters with only vague physical interpretations, and whose computation is time-consuming and not robust.

2. Experimental

We have been using a MOSESII eNose [9] with two sensor modules: an eight-sensor QMB module, and an eight-sensor MOX module. (Reviews on these technologies can be found in, e.g., [1,10].) The samples were put in 20 ml vials in HP7694 headspace sampler, which heated them to 40 °C and injected the headspace content into the electronic nose. There, the analyte was first introduced into the QMB chamber, whence it followed to the 300 °C heated MOX chamber. The injection lasts for 30 s, and is followed by a 15 min purging stage using synthetic air.

We have tested our models against a large dataset, composed of 30 volatile odorous pure chemicals listed alphabetically in Table 1. These chemicals were intentionally chosen from many different families, so that they would represent a broad range of possible stimuli. Each chemical was measured in batches, with a single batch containing at least seven successive measurements. Different batches of the same chemical were usually taken in totally different dates. In total, we have performed 300 measurements, with an average of 10 per chemical.

Table 1
The 30 pure chemicals in our dataset

List of chemicals	
1*S*-(−)-α-pinene	Ethyl-2-methylbutyrate
1*S*-(−)-β-pinene	Ethyl-3-methylthiopropionate
1-Phenyl-1,2-propanedione	Ethyl-*n*-valerate
2-Acetylpyridine	Ethyl acetoacetate
2,3-Heptanedione	Ethyl caproate
4-Methylanisole	Ethylpyrazine
α-Angelica lactone	Phenylacetaldehyde dimethyl acetal
Amyl butyrate	Propylidene phthalide
Butyl butyrate	*R*-(−)-limonene
Butyl butyryl lactate	*S*-(−)-limonene
Butylidene phthalide	Terpinotene
Cis-3-hexenyl acetate	*Trans*-2-hexenal
Cis-6-nonenal	*Trans*-2-hexenol
Citral	*Trans*-2-methyl-2-pentenoic
Ethyl-2-methyl-4-pentenoate	*Trans*-2-octenal

3. Standard feature extraction methods

The signals obtained from eNoses normally have one of two basic shapes, distinguished only by the duration of the stimulus presentation. When the stimulus is introduced long enough for the sensor to reach a steady state (typically a couple of minutes), a *steady-state signal* of the shape shown in Fig. 1b is obtained. But, when the stimulus is introduced only for a short duration (typically 20–30 s), a *transient signal* is obtained, of the general form shown in Fig. 1a.

The common practice in the field is to represent each signal using a single feature, having some simple, purely geometric, definition. Here are some examples of the most popular of these.

Let $\psi_i(t)$ be the measured signal of the ith sensor. If it is a steady-state signal, the custom is to choose the feature as the difference between the steady-state response and the base-

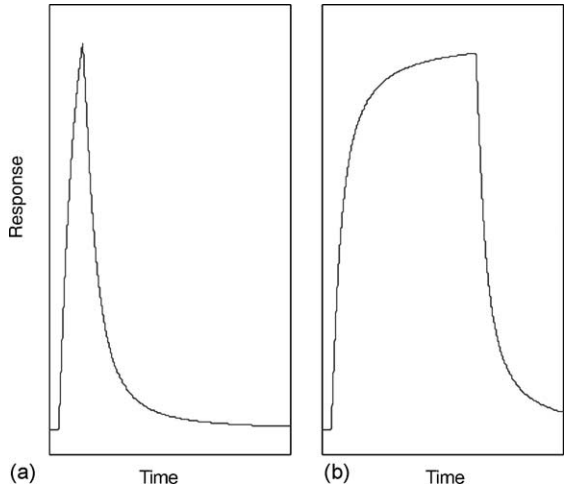

Fig. 1. Typical shapes of eNoses signals: (a) transient signal: the stimulus is introduced for a relatively short time; (b) steady-state signal: the stimulus is introduced for a relatively long time.

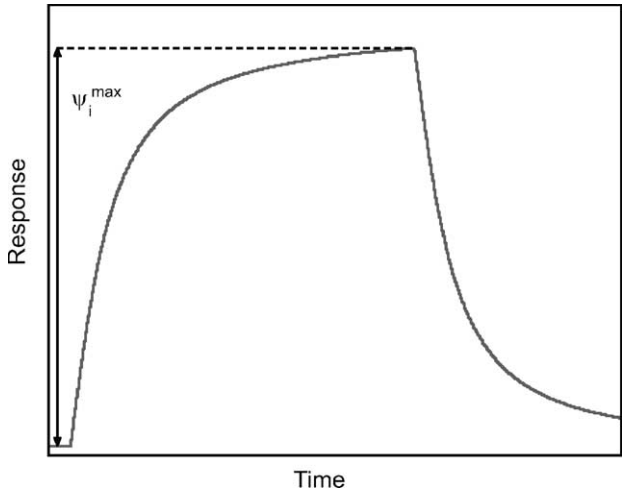

Fig. 2. Definition of $\psi_i^{\max}$ for a steady-state signal.

line, $\psi_i^{\max}$; see Fig. 2. For transient signals, the repertoire of features is richer; see Fig. 3. The most popular feature is, again, the difference between the signal's peak and its baseline, $\psi_i^{\max}$ (Fig. 3a). Other options are to take the area beneath the curve, A_i (Fig. 3b), the area beneath the curve left of the peak, $A_i^{\max}$ (Fig. 3c), and the time it takes for the signal to reach its peak, $T_i^{\max}$ (Fig. 3d). In several cases, more than one feature per signal is used, working with a subset of the aforementioned features.

While these methods have the advantage of being simple and fast to compute, their primary weakness is in their purely geometrical nature. They do not take into consideration any specific properties of the sensors, thus leaving some of the potential information carried in the signals unutilized. These features have been successfully used for many simple applications, but it is generally agreed that more sophisticated features will be required when turning to more demanding tasks.

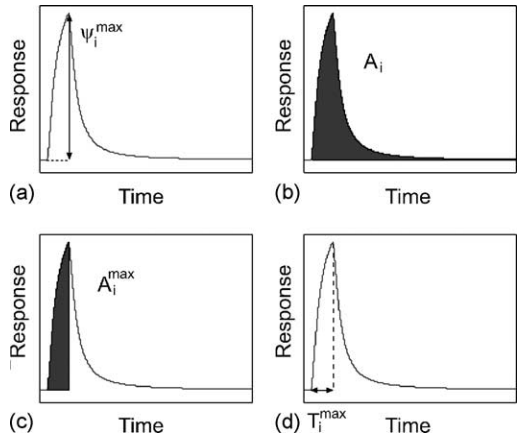

Fig. 3. Definition of the four most popular features in transient signals: (a) the difference between the peak and the baseline, $\psi_i^{\max}$; (b) the area under the curve, A_i; (c) the area under the curve left of the peak, $A_i^{\max}$; (d) the time from the beginning of the signal to the peak, $T_i^{\max}$.

Using different feature extraction methods is rare. For example, White et al. [11] built a special artificial neural network that is fed the entire signal. Their method uses, of course, the maximum possible information, but taking the entire signal seems superfluous and computationally expensive. Also, the neural network is highly complex and appears to be hard to implement.

4. Feature extraction based upon response models

In this section we present our analytic models of a sensor's response, and the features associated with them. First, we develop a simple model of the apparatus, from which we obtain the Lorentzian and the exponential models. We then introduce the double-sigmoid empirical model, and give some heuristics regarding it. Quantitative analysis of the performance of each model is postponed to Section 5. In the experiments that we have been carrying out, we used only short duration injection (30 s). Consequently, we have developed and tested our models only against transient signals; see Fig. 1a. Nevertheless, we anticipate that our models will also fit steady-state signals.

4.1. The Lorentzian and exponential models

We assume that the measurement system is composed of n dimensionless sensors, arranged in succession in a chamber, and that the flow of particles through the chamber is one-dimensional, along the x-axis, as described in Fig. 4. Let x_i be the coordinate of the ith sensor, and let $x = 0$ be the chamber's inlet. Let $N_p(x, t)$ be the number of particles of the inspected chemical in location x at time t, and let $f(t)$ be the number of particles at the inlet at time t. By definition

$$N_p(0, t) = f(t), \tag{1}$$

$f(t)$ is still unspecified, but it must have the general property of being non-negative everywhere. Moreover, if we use N_0 to denote the total number of particles introduced into the chamber during the measurement, then

$$\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = N_0. \tag{2}$$

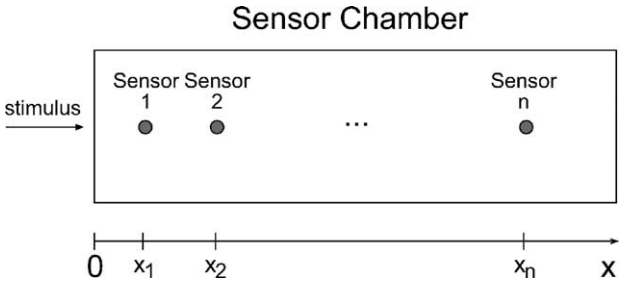

Fig. 4. A schematic description of a one-dimensional sensor chamber.

Next, let us assume that all the particles move in the chamber along the x-direction at a constant velocity v. The profile of the particles near the ith sensor is

$$N_p(x_i,t) = N_p\left(0, t-\frac{x_i}{v}\right) = N_p(0,t-t_i) = f(t-t_i), \qquad (3)$$

where $t_i = x_i/v$ is just the time it takes for a particle to make its way between the inlet and sensor i. In deriving (3) we implicitly assume that only a small fraction of the particles interact with the sensors, so that the number of particles that drop out of the stream is negligible. A sensor, then, is not influenced by the fact that there are other sensors preceding it.

Let $N_i(t)$ be the number of particles that are adsorbed at the ith sensor at time t. We assume that $N_i(t)$ is a constant percentage k_i of the total number of particles in the vicinity of the sensor, so that

$$N_i(t) = k_i N_p(x_i,t) = k_i f(t-t_i). \qquad (4)$$

Let $g_i(t_2-t_1)$ be the probability that a particle that was adsorbed into the ith sensor at time t_1, is still present on it at time t_2. From physical considerations, $g_i(t)$ must be monotonically decreasing, and must obey

$$g_i(0) = 1, \qquad \lim_{t\to\infty} g_i(t) = 0.$$

Let $L_i(t)$ be the total number of particles that are present on the ith sensor at time t. In terms of $g(t)$ and $f(t)$, $L_i(t)$ is

$$\begin{aligned} L_i(t) &= \int_0^\infty N_i(t-u)g_i(u)\,\mathrm{d}u \\ &= k_i \int_0^\infty g_i(u)f(t-t_i-u)\,\mathrm{d}u. \end{aligned} \qquad (5)$$

Let $R_i(t)$ be the response of the ith sensor at time t. Assuming that $R_i(t) = a_i L_i(t)$, with a_i a sensor-specific constant, we get

$$R_i(t) = \alpha_i \int_0^\infty g_i(u)f(t-t_i-u)\,\mathrm{d}u, \qquad (6)$$

where $\alpha_i = a_i k_i$. This is the most general form of the response. However, we can bring it to a different form if we use the fact that stimulus presentation is always bounded in duration. Accordingly, we may take $f(t)$ to be non-zero only in the range $t \in [0,T]$, arbitrarily setting $t=0$ at the time of the beginning of the stimulus presentation. Using this fact, and introducing the new variable $v = t-t_i-u$, we get

$$R_i(t) = \begin{cases} 0, & t < t_i, \\ \alpha_i \int_0^{t-t_i} g_i(t-t_i-v)f(v)\,\mathrm{d}v, & t_i \le t \le t_i+T, \\ \alpha_i \int_0^{T} g_i(t-t_i-v)f(v)\,\mathrm{d}v, & t > t_i+T. \end{cases} \qquad (7)$$

Even without specifying $f(t)$ and $g(t)$, we can point out some interesting properties of (7):

(1) For $t > t_i+T$, the function $R_i(t)$ is monotonically decreasing. To prove this, we simply have to differentiate the appropriate expression in (7),

$$\frac{\mathrm{d}R_i(t)}{\mathrm{d}t} = \alpha_i \int_0^T \frac{\mathrm{d}g_i(t-t_i-v)}{\mathrm{d}t} f(v)\,\mathrm{d}v, \quad t > t_i+T.$$

But $g_i(t)$ is known to be monotonically decreasing, while $f(t)$ in this range is positive. Thus, $(\mathrm{d}g_i/\mathrm{d}t)f$ is negative, resulting in negative $\mathrm{d}R_i/\mathrm{d}t$. For a measured signal with a peak at $t_i + T_i^{\max}$, we are thus guaranteed that $T \ge T_i^{\max}$.

(2) Let us look at large times $t \gg t_i+T$. If the function $g(t)$ is smooth enough at these times, which is reasonable, we can approximate the third expression in (7) by

$$\begin{aligned} &R_i(t) \approx \alpha_i g_i(t) \int_0^T f(v)\,\mathrm{d}v = \alpha_i N_0 g_i(t), \\ &t \gg t_i+T. \end{aligned} \qquad (8)$$

Therefore, examination of the time-dependency of the far end of the signal can give us some information about the shape of $g(t)$.

4.1.1. Evaluating $f(t)$

The function $f(t)$ captures the shape of the injected stimulus over time. Since the injection time in our system (30 s) is small relative to the total measurement time (600 s), we speculate that the actual shape of $f(t)$ is mostly relevant in the short period of the fast rise of the signal, and has a much smaller impact on the dominant decreasing part. Consequently, we do not expect the model to be very sensitive to the choice of $f(t)$. Since the injection is controlled by the headspace autosampler, it is assumed to be quite homogeneous in time, and it is thus easiest to assume that the stimulus is injected at a constant rate:

$$f(t) = \begin{cases} \dfrac{N_0}{T}, & 0 \le t \le T, \\ 0 & \text{otherwise.} \end{cases}$$

Substituting this in (7), we get

$$R_i(t) = \begin{cases} 0, & t < t_i, \\ \beta_i \int_0^{t-t_i} g_i(v)\,\mathrm{d}v, & t_i \le t \le t_i+T, \\ \beta_i \int_{t-t_i-T}^{t-t_i} g_i(v)\,\mathrm{d}v, & t > t_i+T, \end{cases} \qquad (9)$$

where $\beta_i = N_0\alpha_i/T$. For this choice of $f(t)$, we can prove that the signal's peak occurs exactly at time t_i+T. To see this, let us differentiate the intermediate term of (9)

$$\frac{\mathrm{d}R_i(t)}{\mathrm{d}t} = \beta_i g_i(t-t_i) \ge 0, \quad t_i \le t \le t_i+T,$$

so that the derivative is always positive. Combining this with our previous result that the derivative is always negative for $t > t_i+T$, we conclude that the peak occurs exactly at t_i+T.

4.1.2. Evaluating $g(t)$

It is tempting to assume an exponential decay, with $g_i(t) = e^{-t/\tau_i}$.

For many physical systems—perhaps the most famous of which is the concentration decay of a radioactive source—this function is a natural choice. Substituting in (9), we obtain a response function of the form

$$R_i(t) = \begin{cases} 0, & t < t_i, \\ \beta_i \tau_i (1 - e^{-[(t-t_i)/\tau_i]}), & t_i \le t \le t_i + T, \\ \beta_i \tau_i (e^{T/\tau_i} - 1)\, e^{-[(t-t_i)/\tau_i]}, & t > t_i + T. \end{cases} \quad (10)$$

This response model is the one we call the *exponential model*. An example of how well it fits a measured signal is shown in Fig. 5a. Although faithfully following the general shape of the signal, the model fails to describe the exact form of the decreasing part. This problem appeared in all the signals that we have been checking, and motivated us to come up with a better model.

In looking for a better function $g(t)$, we scanned many potential candidates against each and every signal in our dataset. Careful analysis of the results isolated one function that stood above all others in its ability to explain the signal shape. This is the Lorentzian decay function:

$$g_i(t) = \frac{\tau_i^2}{t^2 + \tau_i^2}. \quad (11)$$

Substituting it in (9), we obtain a response of the form

$$R_i(t) = \begin{cases} 0, & t < t_i, \\ \beta_i \tau_i \tan^{-1}\left(\dfrac{t - t_i}{\tau_i}\right), & t_i \le t \le t_i + T, \\ \beta_i \tau_i \left[\tan^{-1}\left(\dfrac{t - t_i}{\tau_i}\right) - \tan^{-1}\left(\dfrac{t - t_i - T}{\tau_i}\right)\right], & t > t_i + T. \end{cases} \quad (12)$$

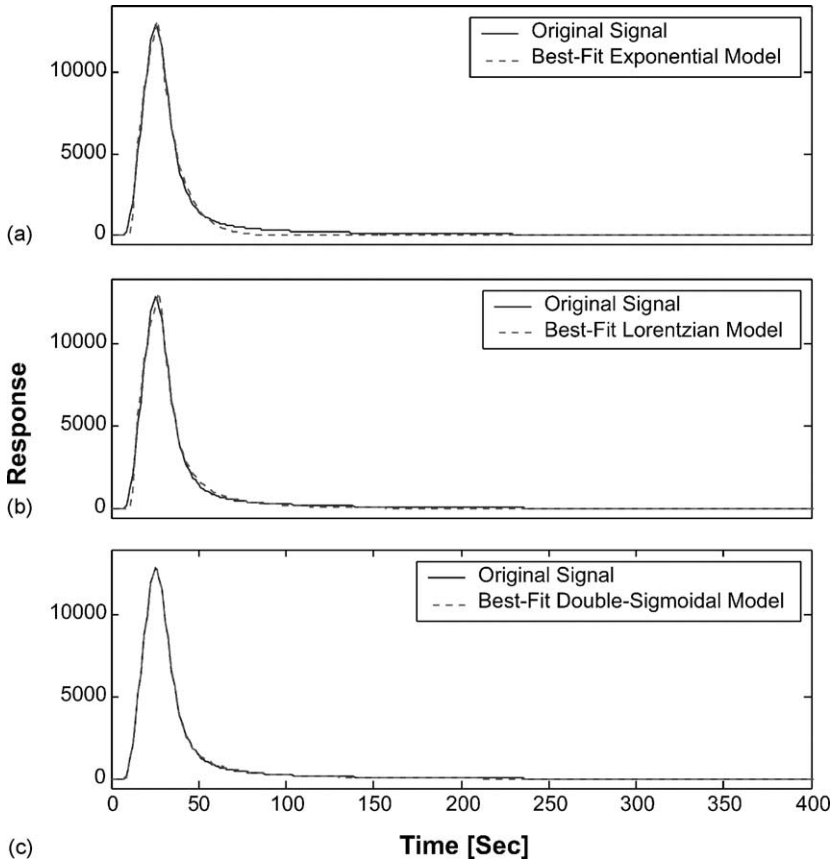

Fig. 5. A comparison between a typical signal (*cis*-3-hexenyl acetate measured with a QMB sensor) and the three analytic models: (a) comparison with the exponential model: notice the deviation in the decreasing part region; (b) comparison with the Lorentzian model: a vast improvement in the global fit is observed; (c) comparison with the double-sigmoid model: the fit is almost perfect, the differences between the measured signal and the model are indistinguishable.

This is the explicit form of our *Lorentzian model*. This time, the decreasing part, as well as the peak region, are nicely captured; see the example in Fig. 5b.

4.1.3. Implementation

Since neither of the models is everywhere differentiable, we could not use gradient-based methods for the curve fitting, and preferred the Matlab® function *fminsearch*, which uses the simplex search method [12].

The speed of convergence dramatically depends on the initial guess of the parameters. Luckily, all parameters are physically meaningful for both the exponential and the Lorentzian models, so that we are able to supply an excellent initial guess, as follows:

(1) t_i is just the time when the signal starts to rise.
(2) T is just $T_i^{\max}$.
(3) τ_i characterizes the decay time of the signal, which we have found not to fluctuate too much for different stimuli. Examining the entire dataset, we found a typical value of τ_i for each sensor, that is used for initialization. The values are listed in Table 2.
(4) β_i is related to the amplitude of the signal. From (10) and (12) the value of the signal at the peak is $\psi_i^{\max} = \beta_i \tau_i \tan^{-1}(T/\tau_i)$ for the Lorentzian model, and $\psi_i^{\max} = \beta_i \tau_i (1 - \exp(-T/\tau_i))$ for the exponential model, so that our initial guess for β_i is

$$\beta_i = \begin{cases} \dfrac{\psi_i^{\max}}{\tau_i \tan^{-1}(T/\tau_i)} & \text{for the Lorentzian model,} \\ \dfrac{\psi_i^{\max}}{\tau_i (1 - \exp(-T/\tau_i))} & \text{for the exponential model.} \end{cases}$$

4.2. The double-sigmoid model

Developing an analytic expression based on a model of the system is a possible approach. Another possibility is to check measured signals against a large library of candidate models. We have used a library of 417 candidates, of which some are classical asymmetric peak functions (such as log-normal, extreme value, and Gamma function), but the majority are peak functions that we have constructed by multiplying two opposing sigmoid functions (one monotonically decreasing and the other monotonically increasing). Testing each of the candidates against our dataset, one function outperformed all the others in its ability to fit the measured signals, both of QMB sensors and of MOX sensors. This nine parameter function, which is the one that we call the double-sigmoid model, looks like this:

$$R_i(t) = \frac{\alpha_i}{\pi}\left[1 - \exp\left(-\left(\frac{t-\beta_i}{\gamma_i} + \epsilon_i\right)^{\delta_i}\right)\right]^{\eta_i} \times \left[\frac{\pi}{2} - \tan^{-1}\left(\frac{t-\mu_i}{\nu_i}\right)\right]^{\lambda_i} \tag{13}$$

with α_i, β_i, γ_i, δ_i, ϵ_i, η_i, μ_i, ν_i, and λ_i the free parameters. $R_i(t)$ yields an impressive fit with the measured signals, being practically indistinguishable in most of the cases. An example of a fit is shown in Fig. 5c. The function in the left-hand square brackets is the sigmoid describing the rising part of the signal. It is an hybridization of two familiar cumulative probability distribution functions: for $\epsilon_i = (\ln 2)^{1/\delta_i}$ it is the cumulative Weibull probability distribution function

$$1 - \exp\left(-\left(\frac{t-\beta_i}{\gamma_i} + (\ln 2)^{1/\delta_i}\right)^{\delta_i}\right),$$

while for $\delta_i = 1$, $\epsilon_i = -\ln(1 - \sqrt{2}/2)$, and $\eta_i = 2$, it is the pulse cumulative probability distribution function

$$1 - \exp\left(-\left(\frac{t-\beta_i}{\gamma_i} - \ln\left(1 - \frac{\sqrt{2}}{2}\right)\right)\right)^2.$$

See, e.g., in TableCurve 2D®, the list of built-in functions. The function in the right-hand square brackets is proportional to 1 − the cumulative Lorentzian probability distribution function

$$\frac{1}{\pi}\left[\frac{\pi}{2} + \tan^{-1}\left(\frac{t-\mu_i}{\nu_i}\right)\right].$$

Here again we have revealed a connection between the decreasing part of the signal and the Lorentzian distribution, thus strongly supporting the foundations of the Lorentzian model.

4.2.1. Implementation

We used both Matlab® optimization functions *fminsearch* and *lsqcurvefit*. The value of the parameters strongly depends on their initial guess; however, lacking a clear physical interpretation, a good guess is difficult to achieve.

As a general rule, α_i is a measure of the signal's amplitude, β_i and γ_i are related to the center and width of the rising sigmoid, respectively, and μ_i and ν_i are related to the center and width of the decreasing sigmoid, respectively. η_i and λ_i are powers normally in the range 0.5–3. Yet, whatever

Table 2
Initialization values for the parameter τ_i for each of the sensors in both the Lorentzian and the exponential models[a]

Model	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
Lorentzian	11	11	10	9	10	18	7	5	93	48	95	80	93	103	90	136
Exponential	13	13	14	12	13	23	9	7	115	60	116	99	111	127	109	159

[a] Here Q_1–Q_8 are the eight QMB sensors, and S_1–S_8 are the eight MOX sensors.

Table 3
Averages and medians of the R^2-test for the curve fitting of our three analytic models[a]

Sensor	Double-sigmoid		Lorentzian		Exponential	
	Average	Median	Average	Median	Average	Median
Q_1	0.99173	0.99975	0.97803	0.99568	0.97124	0.98931
Q_2	0.99926	0.99986	0.99349	0.99601	0.98811	0.99115
Q_3	0.99485	0.99968	0.98664	0.99535	0.9738	0.99182
Q_4	0.99699	0.99971	0.97292	0.99486	0.965	0.98905
Q_5	0.99958	0.99974	0.99421	0.99596	0.98924	0.9918
Q_6	0.99795	0.99959	0.98699	0.99179	0.97239	0.97684
Q_7	0.99944	0.99967	0.99314	0.99611	0.9883	0.99205
Q_8	0.99846	0.99936	0.99204	0.99733	0.98568	0.99201
S_1	0.99923	0.99989	0.98249	0.99222	0.97424	0.98447
S_2	0.9991	0.99969	0.96444	0.97624	0.94582	0.95751
S_3	0.99889	0.99983	0.97726	0.99105	0.96846	0.98188
S_4	0.99949	0.99976	0.97695	0.98344	0.968	0.97332
S_5	0.99956	0.99986	0.98684	0.99324	0.97805	0.98486
S_6	0.99941	0.99985	0.9782	0.9883	0.97118	0.98072
S_7	0.99929	0.99985	0.98379	0.99039	0.97303	0.97941
S_8	0.99764	0.99988	0.97903	0.9952	0.97414	0.99178

[a] Q_1–Q_8 are the eight QMB sensors, and S_1–S_8 the eight MOX sensors. For all sensors, whether QMB or MOX, all three models give a good fit, with the double-sigmoid model exhibiting excellent fit, the Lorentzian model following, and the exponential model being the poorest of the three.

initialization one takes, the convergence is slow, and sometimes ends up with a non-optimal set of parameters, due to convergence to a local minimum.

5. Results

We first compare our three models with respect to goodness-of-fit, computation speed, and robustness, and see why we consider the Lorentzian model to be the most successful of them all. We then compare the Lorentzian and exponential models to standard feature extraction methods and see their advantages.

5.1. Comparing the analytic models

To quantify how well a model fits the data, we used the well known R^2-test [13] to measure the goodness-of-fit. R^2 is bounded from above by 1, and the closer it gets to 1, the better is the fit in the least squares sense. The advantage of the R^2-test is that it measures goodness-of-fit on a normalized scale, thus enabling comparison between differently scaled signals. We tested our three models against all 300×8 QMB signals, and 300×8 MOX signals, and calculated the average and the median of R^2. The results are shown in Table 3, nicely demonstrating our claims that the double-sigmoid model fits the data extremely well, that the Lorentzian model is next in performance, and that the exponential model is the poorest of the three but still yields pretty satisfying fits.

Moreover, the three models significantly differ in computation time and in robustness, with the Lorentzian and exponential models quite alike, and much better than the double-sigmoid model. For the former, using a Matlab® non-optimized code, a typical 16-signal measurement is processed within a few seconds,[1] while for the latter the computation lasts a few minutes. As for what we call robustness, while the 'correct' set of parameters is always achieved for the Lorentzian and exponential models, convergence to an 'improper' set of parameters is sometimes detected for the double-sigmoid model, implying the existence of substantial local minima in the curve-fitting minimization problem.

Taking into consideration all the above factors, our inevitable conclusion is that the Lorentzian model is the one to be preferred for general data analysis. Yet, we do not rule out the possibility that the exponential model will turn out to better fit sensor types not tested by us, or that the double-sigmoid model will be favored for certain specific small datasets.

5.2. Feature extraction for classification tasks

Data analysis never ends with extracting the features, and the true evaluation of the features lies in how well they serve in subsequent algorithms. Stimuli classification is by far the most popular application of eNoses, and therefore we have decided to test the degree of usefulness of the different features to classification tasks.

Classification can be viewed as a well-studied application of algorithms in the general area of pattern recognition [2]. Many of these algorithms normally require a preliminary learning phase, in which they are trained in classifying measurements of a specific dataset. The learning phase uses a set of measurements—the *training set*—whose actual class

[1] Using C code instead of Matlab® is expected to significantly accelerate the computation time, making it completely negligible.

associations are known in advance. When this phase is finished, the algorithm is ready to be applied to measurements whose class association is not known. It is customary to evaluate the performance of a classification algorithm by applying it on another set of measurements—the *test set*. These are measurements for which we do know the class associations, but we apply the algorithm to them as if we did not. Thus we can compare the classification predicted by the algorithm with the actual one. The tests that we have run consist of the following elements:

- *Classification schemes*: Nine classification schemes, based upon two kinds of algorithms. A k-nearest-neighbors (k-NN) algorithm [2] with $k = 1, 3, 5, 7$, and classification by the shortest Mahalanobis distance [2] in the principal components space of dimensions 1–5.
- *Training set versus test set*: Two alternative ways to divide our 30-chemical dataset into a training set and a test set. (a) The *excess dataset*, generated by taking the first seven measurements of each chemical to be in the training set, and taking any additional measurement to be in the test set (thus having 30% of the measurements in the test set). (b) The *random dataset*, generated by choosing at random 40% of the measurements to be in the test set, leaving a training set of 60% of the data.
- *Features*: When having more than a single feature per signal one gains some degrees of freedom in choosing the best subset of features. If, for example, a fit of a certain signal to the Lorentzian model gives the four parameters β^{Lor}, τ^{Lor}, T^{Lor}, and t^{Lor}, then we could choose as our feature set any of the 15 possible non-empty subsets of these four. We tested the performance of each of these 15 feature sets for both the Lorentzian and the exponential models. For comparison, we also tested the performance of the four standard features $\psi^{\max}$, A, $A^{\max}$ and $T^{\max}$. Furthermore, we have defined a more complex use of a feature set, to be denoted by a *maj* prefix. maj(feature$_1$, feature$_2$, ..., feature$_n$) means the following: use each of the n features separately for the purpose of classification, and then decide on the final classification by a majority rule. In total we tested 51 feature sets listed in Table 4, including some 17 promising majority sets.

For each of the 51 feature sets we applied the nine classification schemes to 101 datasets—one excess dataset and 100 random ones. Performance, measured as the average classification rate, was calculated for each of the feature sets. The results are shown in Table 4, sorted from the best feature set to the worst. As can be seen, the best feature set, which is the one that we recommend to use in general, is $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$, meaning 'classify by majority rule from among $\psi^{\max}$ and the parameters of the Lorentzian model'. For comparison, the set $\mathrm{maj}(\psi^{\max}, A, A^{\max}, T^{\max})$ (i.e., taking majority rule from among all standard features) is only

Table 4
Ranking of all feature sets studied[a]

Rank	Feature set	Rank	Feature set
1	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	27	(β^{Lor})
2	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	28	$(\beta^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$
3	$\mathrm{maj}(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}})$	29	$(\beta^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$
4	$\mathrm{maj}(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	30	$(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}})$
5	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}})$	31	$(\tau^{\mathrm{Exp}}, T^{\mathrm{Exp}})$
6	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	32	$\mathrm{maj}(\psi^{\max}, A, T^{\max})$
7	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}})$	33	(β^{Exp})
8	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	34	(τ^{Lor})
9	$\mathrm{maj}(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	35	$(\tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$
10	$\mathrm{maj}(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	36	$(\psi^{\max})$
11	$(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	37	(τ^{Exp})
12	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}})$	38	$(\tau^{\mathrm{Lor}}, t^{\mathrm{Lor}})$
13	$(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	39	$(\tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$
14	$\mathrm{maj}(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	40	$(\beta^{\mathrm{Exp}}, t^{\mathrm{Exp}})$
15	$(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}})$	41	$(\tau^{\mathrm{Exp}}, t^{\mathrm{Exp}})$
16	$(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}})$	42	$(A^{\max})$
17	$\mathrm{maj}(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}})$	43	$(t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$
18	$(\tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	44	$(t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$
19	$(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	45	(T^{Exp})
20	$(\beta^{\mathrm{Exp}}, T^{\mathrm{Exp}})$	46	$(\beta^{\mathrm{Lor}}, t^{\mathrm{Lor}})$
21	$\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}})$	47	(A)
22	$(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$	48	(T^{Lor})
23	$\mathrm{maj}(\psi^{\max}, A, T^{\max})$	49	$(T^{\max})$
24	$(\beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}})$	50	(t^{Exp})
25	$\mathrm{maj}(\psi^{\max}, A, A^{\max}, T^{\max})$	51	(t^{Lor})
26	$(\beta^{\mathrm{Lor}}, T^{\mathrm{Lor}})$		

[a] The lower the rank the higher the classification rate associated with the set.

in the 25th place, the set $(\psi^{\max})$ is in the 36th place, and the sets (A), $(A^{\max})$, and $(T^{\max})$ are in the 47th, 42nd, and 49th places, respectively. The best non-majority feature set is $(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$. Comparing the rankings of the Lorentzian and exponential models, we see that the Lorentzian model is better also with respect to classification rate. Obviously, they are both much better than the standard features.

It is not possible to give absolute classification performance in Table 4, since the ranking is based on many different classification schemes and datasets. However, to get some impression of the actual classification rates achieved, here are several examples:

(1) Classification by the k-NN algorithm (with $k = 3$) on the excess dataset of the eight QMB sensors:
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 97.8% correct classification.
   - $(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 90% correct classification.
   - $(\psi^{\max})$ gives 84.4% correct classification.
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$ gives 96.7% correct classification.

(2) Classification by shortest Mahalanobis distance in the four-dimensional space of the first four principal components, on the excess dataset of the eight QMB sensors:
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 83.3% correct classification.
   - $(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 96.7% correct classification.
   - $(\psi^{\max})$ gives 42.2% correct classification.
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$ gives 60% correct classification.

(3) Classification by the k-NN algorithm (with $k = 5$) on the random dataset of the eight MOX sensors:
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 93.1% correct classification.
   - $(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, T^{\mathrm{Lor}})$ gives 90% correct classification.
   - $(\psi^{\max})$ gives 86.9% correct classification.
   - $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Exp}}, \tau^{\mathrm{Exp}}, t^{\mathrm{Exp}}, T^{\mathrm{Exp}})$ gives 93.5% correct classification.

6. Summary and discussion

We have introduced a new feature extraction method based upon the idea of fitting measured signals to an analytic model. Among the three models that we have developed, the Lorentzian model (12) was found to be the most powerful, combining fast computation of the parameters, excellent robustness, and good fits to all signals. We found that the most potent set of features was $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$, enabling higher classification rates than other combinations of features.

Besides carrying far more information about the signals, there is another important advantage in using an analytic model. The sensors embedded in an eNose always interact within a certain dynamic range. When an application requires measuring stimuli with different characteristics under the same working conditions, it may happen that it would not be possible to set a dynamic range fitting all stimuli. This would normally be reflected in driving some of the sensors into saturation or even in partial failure. While the computation of standard features necessitates the integrity of the signals, curve-fitting can make do with only parts of the signals—the non-corrupted ones. Therefore, not only can the Lorentzian model parameters be computed even when parts of the signal are corrupted, but they can actually be used to reconstruct the damaged parts. Obviously, in such cases we will not use the feature set $\mathrm{maj}(\psi^{\max}, \beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$, but will rather replace it with $\mathrm{maj}(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}})$ or $\mathrm{maj}(\beta^{\mathrm{Lor}}, \tau^{\mathrm{Lor}}, t^{\mathrm{Lor}}, T^{\mathrm{Lor}})$.

The former claim applies also to the purpose of shortening measurement time. The analytic model can be computed based upon the initial part of the signal, so that one should not wait until the completion of the measurement to extract the features.

One might think that a model-based feature extraction method would be sensitive to the sensor's type and to the apparatus' setup. However, this is not the case. As is evident from Table 3, the model fits equally well two different types of sensor modules—the QMB and the MOX modules. Based on these results, we believe that there are good chances that our model will also fit additional sensor technologies and embeddings, as well as different analytic devices housing chemical sensors.

We have limited our dataset to include only transient signals. The validity of our models to steady-state signals has not yet been checked.

References

[1] J.W. Gardner, P.N. Bartlett, Electronic Noses, Principles and Applications, Oxford University Press, New York, 1999.
[2] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley, New York, 2000.
[3] S. Hahn, M. Frank, U. Weimar, Rancidity investigation on olive oil: a comparison of multiple headspace analysis using an electronic nose and GC/MS, in: Proceedings of the Seventh International Symposium on Olfaction and Electronic Nose, ISOEN 2000, July 2000, pp. 49–50.
[4] C. Di Natale, A. Macagnano, S. Nardis, R. Paolesse, C. Falconi, E. Proietti, P. Siciliano, R. Rella, A. Taurino, A. D'Amico, Comparison and integration of arrays of quartz resonators and metal-oxide semiconductor chemoresistors in the quality evaluation of olive oils, Sens. Actuators B 78 (2001) 303–309.
[5] C. Di Natale, G. Olafsdottir, S. Einarsson, E. Martinelli, R. Paolesse, A. D'Amico, Comparison and integration of different electronic noses for freshness evaluation of cod-fish fillets, Sens. Actuators B 77 (2001) 572–578.
[6] P. Boilot, E.L. Hines, S. John, J. Mitchell, F. Lopez, J.W. Gardner, E. Llobet, M. Hero, C. Fink, M.A. Gongora, Detection of bacteria causing eye infections using a neural network based electronic nose system, in: J.W. Gardner, K.C. Persaud (Eds.), Electronic Noses and Olfaction 2000, IoP Publishing, Bristol, February 2001, pp. 189–196.
[7] Y.-J. Lin, H.-R. Guo, Y.-H. Chang, M.-T. Kao, H.-H. Wang, R.-I. Hong, Application of the electronic nose for uremia diagnosis, Sens. Actuators B 76 (2001) 177–180.

[8] F. Michael, H. Ulmer, J. Ruiz, P. Visani, U. Weimar, Complementary analytical measurements based upon gas chromatography–mass spectrometry, sensor system and human sensory panel: a case study dealing with packaging materials, Anal. Chim. Acta 431 (2001) 11–29.
[9] J. Mitrovics, H. Ulmer, U. Weimar, W. Gopel, Modular sensor systems for gas sensing and odor monitoring: the MOSES concept, Acc. Chem. Res. 31 (1998) 307–315.
[10] H.T. Nagle, S.S. Schiffman, R. Gutierrez-Osuna, The how and why of electronic noses, IEEE Spectrum (September 1998) 22–34.
[11] J. White, T.A. Dickinson, D.R. Walt, J.S. Kauer, An olfactory neuronal network for vapor recognition in an artificial nose, Biol. Cybernet. 78 (1998) 245–251.
[12] Optimization Toolbox for use with Matlab®, User's Guide, Version 2, Fourth Printing (Release 12), The MathWorks, Inc., 2000.
[13] W.R. Dillon, M. Goldstein, Multivariate Analysis Methods and Applications, Wiley, New York, 1984.

Biographies

L. Carmel received his BSc in physics at Tel-Aviv University, Israel, in 1991, and his MSc degree in physics at the Technion, Israel Institute of Technology, in 1998. He is currently completing his PhD studies in the Department of Computer Science and Applied Mathematics at The Weizmann Institute of Science, Israel. His research deals with materializing odor digitization, transmission and reproduction, and it involves many kinds of mathematics (e.g., multivariate data analysis, statistical pattern recognition), biology (e.g., the sense of smell, receptor repertoires), and chemistry (e.g., electronic noses, chemical sensors).

S. Levy received his MSc degree in organic chemistry at the Hebrew University of Jerusalem, Israel, in 1997. His research included biological sensors, the implementation of biosensor for the detection of explosives (vapor and trace analysis), surface modification and the use of electrochemical and electrogravimetric instruments and their utilization for biosensors applications. Additional expertise includes the chemical analysis and detection of explosives and chemical sensors (electronic noses). He is currently employed at Hewlett-Packard Company, Indigo Division.

D. Lancet is the Ralph and Lois Silver Professor of Human Genomics at the Department of Molecular Genetics of The Weizmann Institute of Science, Rehovot, Israel. He has been Head of the Crown Human Genome Center at Weizmann, since 1998. Prof. Lancet received his BSc degree in chemistry at the Hebrew University of Jerusalem in 1970 and his PhD degree in chemical immunology at The Weizmann Institute in 1978. He headed Weizmann's Department of Membrane Research and Biophysics (1995–1997). Lancet pioneered genome research in Israel, and currently operates Israel's National Laboratory for Genome Infrastructure. His research interests include the genetic basis of the sense of smell and of inherited diseases, formalisms of molecular recognition and computer models for the origins of life. As part of an extensive involvement in the bioinformatics scene, his team has developed GeneCards, a world-known compendium of human genes, and GeneNote, a whole-genome DNA array register. He received, among others, the First Takasago Award of the American Association for Chemoreception Sciences (1986), and the R.H. Wright Award in Olfactory Research (1998). Lancet is a member of the European Molecular Biology Organization since 1996.

D. Harel is the William Sussman Professor of Mathematics at The Weizmann Institute of Science in Israel, and has been Dean of the Faculty of Mathematics and Computer Science there since 1998. He is also co-founder of I-Logix, Inc., Andover, MA. He received his BSc from Bar-Ilan University in 1974, his MSc from Tel-Aviv University in 1976, and his PhD from MIT in 1978. He has worked in several areas of theoretical computer science, including computability, finite model theory and logics of programs, and in recent years has become involved in other areas, including software and systems engineering, visual languages, graph layout, modeling and analysis of biological systems, and smell communication. He is the inventor of statecharts, and co-inventor of live sequence charts (LSCs), and was part of the team that designed the Statemate and Rhapsody tools. He has received a number of awards, including ACMs Karlstrom Outstanding Educator Award in 1992. His latest books are "Dynamic Logic" (with Kozen and Tiuryn), MIT Press, 2000, and "Computers Ltd.: What They Really Can't Do", Oxford, 2000.