Available online at www.sciencedirect.com
SCIENCE DIRECT®
Sensors and Actuators B 106 (2005) 95–100
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

Electronic nose signal restoration—beyond the dynamic range limit

Liran Carmel

Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

Available online 23 July 2004

Abstract

When measuring over-concentrated stimuli, chemical sensors tend to exhibit corrupted time signals, which are normally categorized as missing data. Such a failure of one or more sensors occurs frequently in applications where an eNose is exposed to a diverse repertoire of chemicals. As a rule, missing data are removed from the dataset by leaving a potentially large portion of the original dataset unutilized. Here we propose an algorithm to handle such missing data by utilizing intact regions of corrupted signals to restore the damaged regions. We do so by fitting a parametric model of the sensor response over time to the intact regions, and using the resulting model for the restoration. We show that the restoration is both accurate and consistent, thus allowing for the restored signals to take part in any subsequent data analysis process.

Keywords: Electronic nose; Signal restoration; Missing data; Signal corruption; Signal failure

1. Introduction

Many kinds of chemical sensors are embedded in contemporary electronic noses (eNoses), which utilize a rather broad spectrum of physical principles and technologies [1,2]. Whatever the sensor technology used, the response strength is characterized by a monotonic dependence on the stimulus concentration, thus limiting it to a certain "operational" concentration range $[c_{\min}, c_{\max}]$. If the stimulus concentration is below $c_{\min}$—the *detection threshold*—the response is too weak to carry any discriminatory information. In contrast, if the stimulus concentration is above $c_{\max}$—which we may call the *failure threshold*—the response signal is corrupted and cannot be used for data analysis. Usually, the failure threshold is not related to any physical or chemical properties of the sensor, but is rather an effect of the supporting electronic circuits. Hereinafter we shall use the term *dynamic range* to describe this "operational" range $[c_{\min}, c_{\max}]$.

In applications that expose an eNose to one or more similar chemicals, with limited concentration fluctuations (e.g., an eNose that monitors the quality of raw materials), the sensitivity of the sensors can be tuned such that a typical stimulus safely resides inside the dynamic range. In other applications, however, eNoses are expected to record stimuli that were never presented to them before, or stimuli having diverse chemical properties (e.g., an eNose intended to serve as a sniffer in an odor communication system [3]). In such cases, a reasonable strategy would be to fix each sensor in its most likely working point and to carry out special processing when a stimulus is outside the dynamic range. Despite the fact that such special processing can significantly broaden the scope of the applicability of eNoses by algorithmically extending their dynamic range, this has not yet been reported in the literature. Instead, the common practice is to categorize corrupted signals as missing data, and to remove them from the dataset. This, however, might cause to undesirable (and unnecessary) reduction in the size of the dataset. For example, suppose that a 10-sensor eNose is trained to discriminate between five different stimuli, and that one of the stimuli has four of its signals corrupted. In that case, one can either remove that stimulus from the dataset (removing 20% of the original dataset) thus limiting the eNose to discriminate only between four stimuli, or one can remove the four awkward sensors from the dataset (removing 40% of the original dataset) thus utilizing only 6 of the 10 sensors. Either way, a large portion of the dataset is removed, including many intact signals.

In this paper we suggest an algorithm for "fixing" corrupted signals by restoring their damaged parts. By definition, for stimulus concentrations below $c_{\min}$, it is useless to infer anything from the signal since it carries zero, or nearly zero, information. In sharp contrast, corrupted signals often contain wide intact regions that carry much information. In a previous work [4], we developed an analytic form of

E-mail address: liran.carmel@weizmann.ac.il (L. Carmel).

0925-4005/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2004.05.042

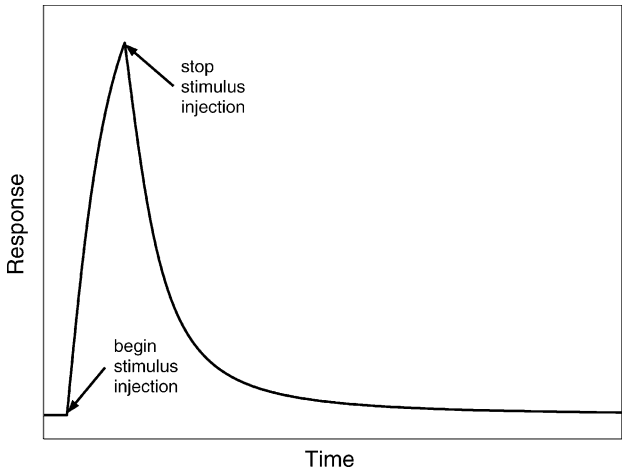

Fig. 1. Schematic shape of a transient signal.

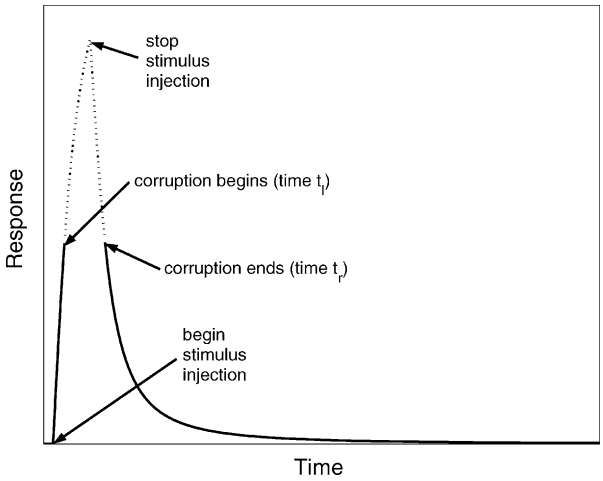

Fig. 2. Dynamics of a failure in a transient signal. The solid lines depict the intact regions of the signal, in between which the corrupted region resides. The dashed line depicts the desired output of a restoration algorithm.

the time-dependent sensor response, characterized by four physically meaningful parameters. Here we show that these parameters can be estimated using only portions of a signal, suggesting the restoration of a corrupted signal based upon its intact regions.

2. Types of corruption

We limit our discussion to transient signals, produced by injecting a stimulus for a relatively short period of time (typically, around 30 s), and then purging the system. Such signals have the typical shape shown in Fig. 1.

Our experiments have shown that when a failure occurs, the sensor recovers when the stimulus concentration drops down again during the purging phase, as is schematically described in Fig. 2. We have further found that a failure appears in one of three forms:

- *Saturated signal*: In the damaged region the sensor shows a constant reading; see Fig. 3a.
- *Raging unbiased signal*: In the damaged region the sensor runs wildly; see Fig. 3b.
- *Raging biased signal*: In the damaged region the sensor runs wildly, and the baseline seems to have been shifted after the recovery; see Fig. 3c and d.

We would like to emphasize that neither the saturation nor the wild running express chemical phenomena on the sensor's surface, but rather are artifacts of the supporting electronics driven outside its dynamic range. Notice, as in Fig. 3b, that the concentration at which the recovery occurs can be different from the concentration at which the corruption commences.

3. The restoration algorithm

In [4] we used a simple physical description of the measurement process to derive an analytic model—the *Lorentzian model*—for the sensor's response over time. Explicitly, the model reads

$$L(t;\theta) = \begin{cases} 0 & t < t_i \\ \beta\tau \tan^{-1}\left(\dfrac{t-t_i}{\tau}\right) & t_i \le t \le t_i + T \\ \beta\tau\left[\tan^{-1}\left(\dfrac{t-t_i}{\tau}\right) - \tan^{-1}\left(\dfrac{t-t_i-T}{\tau}\right)\right] & t > t_i + T \end{cases}$$

where $\theta = \{\beta, \tau, t_i, T\}$ is the set of parameters of the model, t is the time and L is the sensor's response.

Before explaining how to utilize the Lorentzian model for the purpose of signal restoration, we must define some convenient notations. Let us use $s(t)$ to denote the measured time response of a sensor. Let t_l be the time of the last intact point before the corruption commences (l for left), and similarly let t_r be the time of the first intact point after the corruption ceases (r for right); see Fig. 2. When the signal is high above its baseline, the signal-to-noise ratio is also high, resulting in smooth derivatives. Thus, t_l and t_r can be easily found by detecting rapid changes in the derivative. Let us symbolically write $s(t)$ as a composition of two signals, $s(t) = s_I(t) + s_D(t)$, where $s_I(t)$ is the intact region(s) and $s_D(t)$ is the damaged region. More formally,

$$s_I(t) = \begin{cases} s(t) & t \le t_l \text{ or } t \ge t_r \\ 0 & \text{otherwise} \end{cases} \qquad s_D(t) = \begin{cases} s(t) & t_l < t < t_r \\ 0 & \text{otherwise} \end{cases}$$

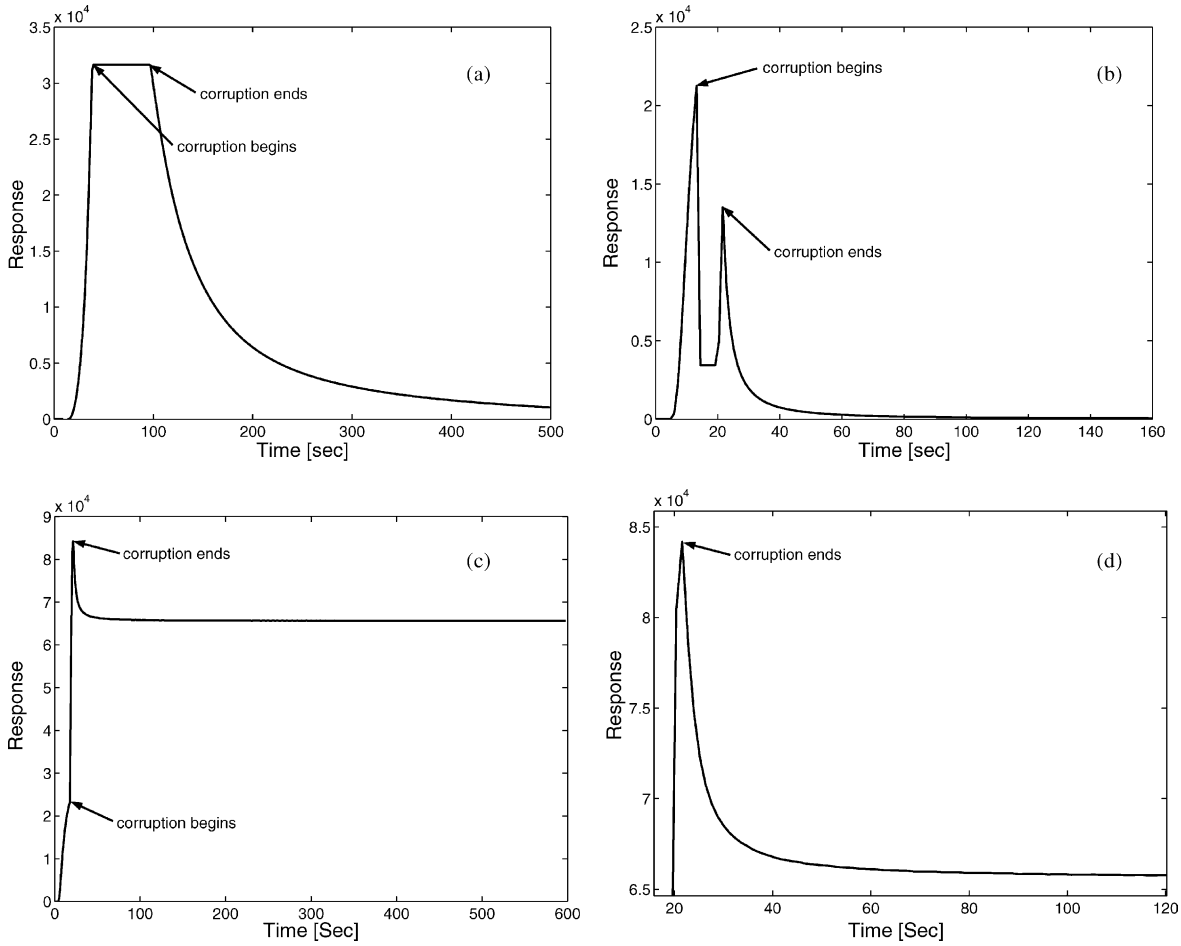

Fig. 3. Examples of the different corruption types: (a) saturated signal, (b) raging unbiased signal, (c) raging biased signal and (d) zoom on the sensor recovery of the raging biased signal shown in (c). Notice that the overall shape of the signal after the recovery seems good despite of the baseline shift.

Analogously, we use $L_{\mathrm{I}}(t;\theta)$ and $L_{\mathrm{D}}(t;\theta)$ for the Lorentzian response calculated at the intact and damaged regions respectively:

$$L_{\mathrm{I}}(t;\theta)=\begin{cases} L(t;\theta) & t\le t_{\mathrm{l}} \text{ or } t\ge t_{\mathrm{r}} \\ 0 & \text{otherwise}\end{cases} \qquad L_{\mathrm{D}}(t;\theta)=\begin{cases} L(t;\theta) & t_{\mathrm{l}}<t<t_{\mathrm{r}} \\ 0 & \text{otherwise}\end{cases}$$

Let θ^* be the estimator of the Lorentzian parameters obtained by curve fitting the Lorentzian model to the intact region $s_{\mathrm{I}}(t)$. A naive restoration algorithm would then be

$$s^*(t)=s_{\mathrm{I}}(t)+L_{\mathrm{D}}(t;\theta^*)$$

where $s^*(t)$ denotes the restored signal. This, however, is not good enough in practice, as can be seen in Fig. 4a. $s(t)$, and even more so $s_{\mathrm{I}}(t)$, obey the Lorentzian model only approximately, and consequently $s^*(t)$ has two undesirable properties: Its height is lower than it should be, and it is not continuous in the boundaries of the damaged region.

To achieve better restoration, we suggest the following iterative algorithm, which forces continuity on $s^*(t)$, and by doing so also increases its height:

1. Correcting for baseline bias: For raging biased signals, as in Fig. 3c, the baseline shift should be corrected by

$$s(t)\leftarrow s(t)-s(\text{last point}) \quad \text{for } t\ge t_{\mathrm{r}}$$

   Here we explicitly assume that the sensor has fully returned to its baseline by the end of the measurement.
2. Modifying $L_{\mathrm{D}}(t;\theta^*)$ to keep $s^*(t)$ continuous: Let $\Delta s_{\mathrm{l}}=s(t_{\mathrm{l}})-L(t_{\mathrm{l}};\theta^*)$ and $\Delta s_{\mathrm{r}}=s(t_{\mathrm{r}})-L(t_r;\theta^*)$ be the deviations of the best-fitting Lorentzian model from the original signal at the boundaries of the damaged region. Let us define a linear correction function

$$\Delta(t)=\Delta s_{\mathrm{l}}+\frac{\Delta s_{\mathrm{r}}-\Delta s_{\mathrm{l}}}{t_{\mathrm{r}}-t_{\mathrm{l}}}(t-t_{\mathrm{l}})$$

   such that $L(t_{\mathrm{r}};\theta^*)+\Delta(t_{\mathrm{r}})=s(t_{\mathrm{r}})$ and $L(t_{\mathrm{l}};\theta^*)+\Delta(t_{\mathrm{l}})=s(t_{\mathrm{l}})$. If we define $\Delta_{\mathrm{D}}(t)$ as

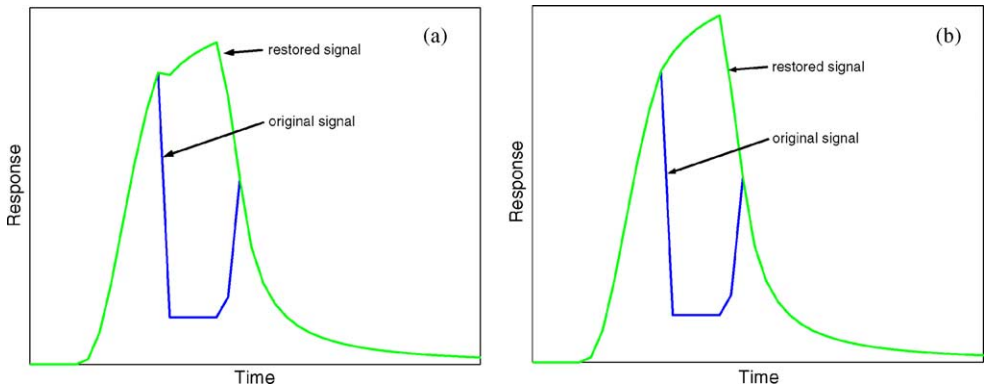

Fig. 4. Comparison between restoration algorithms: (a) the naive algorithm and (b) the iterative algorithm.

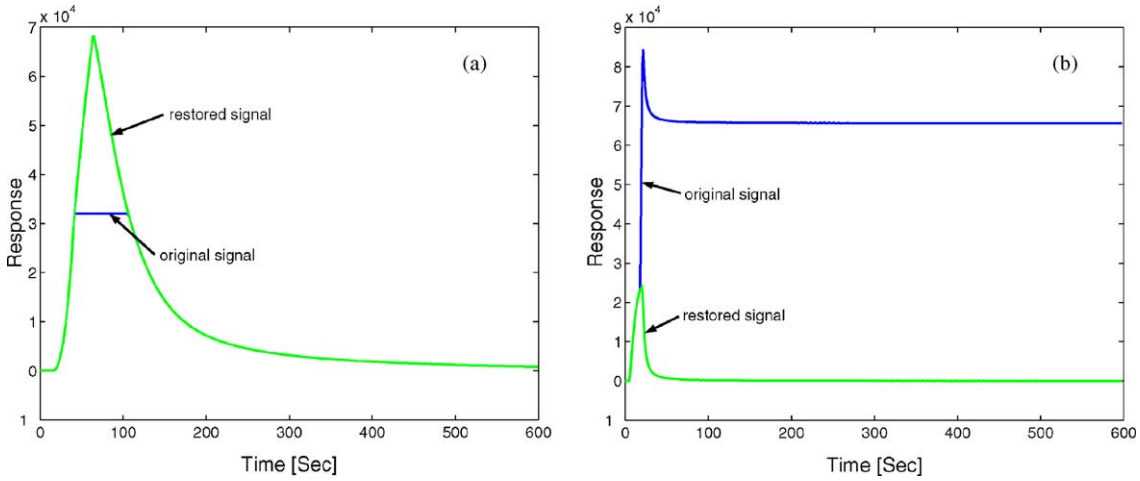

Fig. 5. Restoration at work: (a) saturated signal and (b) raging biased signal.

$$\Delta_{\mathrm{D}}(t) = \begin{cases} \Delta(t) & t_{\mathrm{l}} < t < t_{\mathrm{r}} \\ 0 & \text{otherwise} \end{cases}$$

the restored signal would now be

$$s^*(t) = s_{\mathrm{I}}(t) + L_{\mathrm{D}}(t; \theta^*) + \Delta_{\mathrm{D}}(t)$$

3. Updating θ^*: Find θ^{**}—the estimator of the Lorentzian parameters obtained by fitting the Lorentzian model to $s^*(t)$. If $|\theta^{**} - \theta^*|$ is smaller than some predefined tolerance, stop. Otherwise, substitute $\theta^* = \theta^{**}$ and return to step 2.

Fig. 4b demonstrates the significant improvement in the restoration achieved by the iterative algorithm. Hereinafter, we shall always use the iterative algorithm for restoration. Two additional examples of this restoration, for saturated signal and for raging biased signal, are shown in Fig. 5.

4. Experimental

We have examined our algorithm using data collected by the MOSESII eNose [5] with two sensor modules: an eight-sensor quartz-microbalance (QMB) module, and an eight-sensor metal-oxide (MOX) module. The samples were put in 20-ml vials in an HP7694 headspace sampler, which heated them to 40 °C and injected the headspace content into MOSESII. 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 phase using synthetic air.

5. Results

We have tested two different aspects of our restoration algorithm: How accurate it is, and how much it improves the usability of the dataset. These two issues will be discussed in the next two subsections.

5.1. The accuracy of the restoration

The accuracy of the restoration was measured by artificially clipping uncorrupted signals, and then comparing the original signals with the restored ones. A restoration was defined as successful if the comparison resulted in less

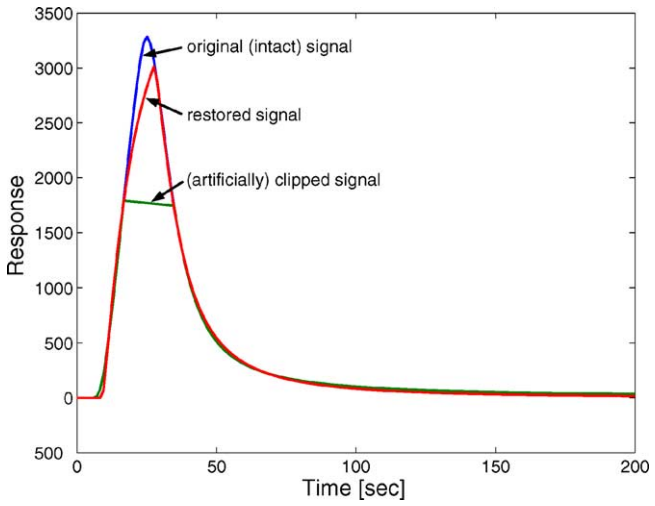

Fig. 6. Comparison between artificially clipped signal (with $c = 50\%$) and its restoration.

than a 10% difference in the signal heights (a typical STD value of MOSESII), and more than 0.99 in the R^2 test for goodness-of-fit. The clipping was performed by removing from the signal all points above c% of the signal's height. After altering c, we concluded that the restoration is successful whenever $c \geq 50\%$. An example of such a restoration for a 50% clipping is shown in Fig. 6.

5.2. The effect on stimuli discrimination

Usually, the accuracy of the restoration is not very important. For classification tasks, which are by far the most popular form of analysis carried out with eNoses, consistency is much more critical. In our context, consistency means that applying the restoration algorithm on two different measurements of the same stimulus should end up with similar restored signals, which in turn should be sufficiently different from the restored signals of another stimulus. In other words, a consistent restoration algorithm should give measurements of the same stimulus clustered together, while keeping clusters of different stimuli sufficiently apart.

To test the consistency of our restoration algorithm, we have used a small (16 measurements) dataset of three chemicals—isoamyl formate, 2,3-hexanedione and 3,4-hexanedione—all having about one-third of their signals corrupted. Using the traditional signal height as a single feature per sensor, we obtain the two-dimensional PCA projection of the raw dataset shown in Fig. 7a. As the signal height is susceptible to corruptions, the different measurements are widely scattered and no clusters can be isolated. However, if we first apply the restoration algorithm on the data, and only then use the signal heights as features, we obtain the two-dimensional PCA projection shown in Fig. 7b. The effect of the restoration, we must say, is overwhelming. Actually, it can be quantified by measuring how well separated are the different clusters in each dataset. To this end let us define the separability index of a dataset as

$$I = \frac{\mathrm{Tr}(S_\mathrm{B})}{\mathrm{Tr}(S_\mathrm{W})}$$

Here S_W is the average within-cluster covariance matrix (the average scatter of a cluster), and S_B is the between-cluster covariance matrix (the scatter of the clusters themselves, where each cluster is represented by its centroid). For more information on these magnitudes see [6]. The reasoning for defining this separability index stems from the fact that the trace of a covariance matrix is proportional to the average Euclidean distance between the cluster members. Thus, the higher is I, the denser the clusters, and better is their separation. The original dataset (before restoration) gives $I = 0.4$, while the restored dataset gives an index higher by two orders of magnitude, $I = 47.5$. This shows that the effect of restoration on the discrimination power of an eNose

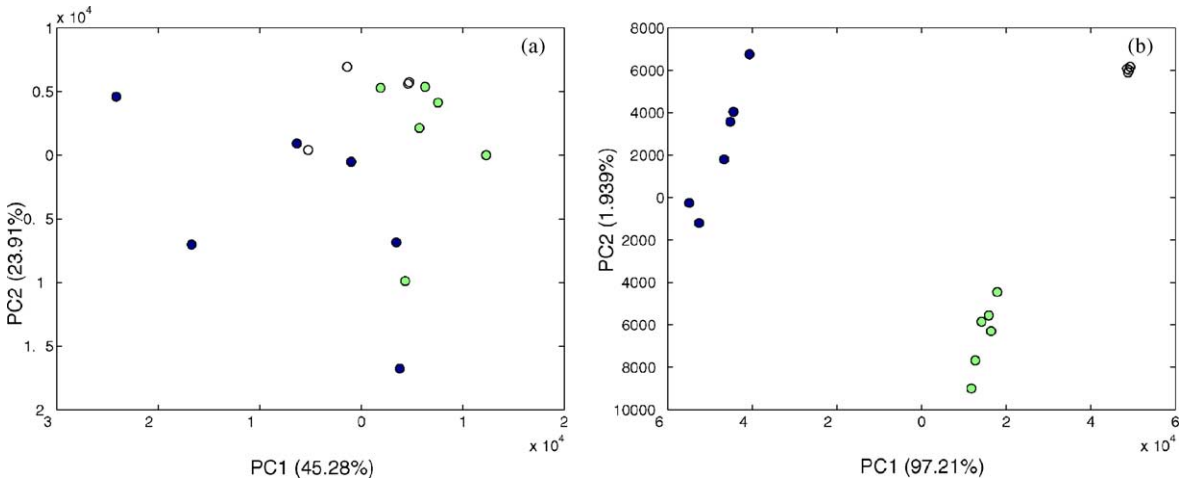

Fig. 7. Two-dimensional PCA projection of (a) the original dataset (before restoration) and (b) the restored dataset. Here, the measurements of isoamyl formate, 2,3-hexanedione and 3,4-hexanedione are indicated by dark, light and hollow dots, respectively.

is striking, as long as, of course, corruption has been taking place.

6. Discussion

As far as we know, this is the first time that an algorithm for handling eNose missing data is proposed. We suggest a way to restore damaged parts in signals that were corrupted due to limited dynamic range of the measurement system. The restoration is shown to be pretty accurate, but more importantly—consistent. Therefore, it enables the usage of samples that originally contained corrupted signals in any data analysis process. This restoration algorithm enhances the capabilities of eNoses, allowing them to deal with broader spectrum of stimuli.

References

[1] J.W. Gardner, P.N. Bartlett, Electronic Noses, Principles and Applications, Oxford University Press, New York, 1999.
[2] T.C. Pears, S.S. Schiffman, H.T. Nagle, J.W. Gardner (Eds.), Handbook of Machine Olfaction: Electronic Nose Technology, Wiley-VCH, Weinheim, 2002.
[3] D. Harel, L. Carmel, D. Lancet, Towards an odor communication system, Comput. Biol. Chem. 27 (2003) 121–133.
[4] L. Carmel, S. Levy, D. Lancet, D. Harel, A feature extraction method for chemical sensors in electronic noses, Sens. Actuators B: Chem. 93 (2003) 67–76.
[5] J. Mitrovics, H. Ulmer, U. Weimar, W. Göpel, Modular sensor systems for gas sensing and odor monitoring: the MOSES concept, Acc. Chem. Res. 31 (1998) 307–315.
[6] A. Webb, Statistical Pattern Recognition, Arnold, London, 1999.