Bioinspired Early Detection through Gas Flow Modulation in Chemo-Sensory Systems

Andrey Ziyatdinov[a,b,*], Jordi Fonollosa[c], Luis Fernández[d,e], Agustín Gutierrez-Gálvez[d,e], Santiago Marco[d,e], Alexandre Perera[a,b]

[a]*Department of ESAII, Universitat Politènica de Catalunya, Pau Gargallo 5, Barcelona, Spain*

[b]*Centro de Investigación Biomèdica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain*

[c]*BioCircuits Institute, University of California, San Diego, La Jolla, CA 92093, USA*

[d]*Signal and Information Processing for Sensing Systems Institute for Bioengineering of Catalonia (IBEC), Baldiri Reixac, 4-8, 08028 Barcelona, Spain*

[e]*Departament dElectrónica, Universitat de Barcelona, Martì i Franquès 1, 08028 Barcelona, Spain*

Abstract

The design of bioinspired systems for chemical sensing is an engaging line of research in machine olfaction. Developments in this line could increase the lifetime and sensitivity of artificial chemo-sensory systems. Such approach is based on the sensory systems known in live organisms, and the resulting developed artificial systems are targeted to reproduce the biological mechanisms to some extent. Sniffing behaviour, sampling odours actively, has been studied recently in neuroscience, and it has been suggested that the respiration frequency is an important parameter of the olfactory system, since the odour perception, especially in complex scenarios such as novel odourants exploration, depends on both the stimulus identity and the sampling method. In this work we propose a chemical sensing system based on an array of 16 metal-oxide gas sensors that we combined with an external mechanical ventilator to simulate the biological respiration cycle. The tested gas classes formed a relatively broad combination of two analytes, acetone and ethanol, in binary mixtures. Two sets of low-frequency and high-frequency features were extracted from the acquired signals to show that the high-frequency features contain information related to the gas class. In addition, such information is available at early stages of the measurement, which could make the technique

---

*Corresponding author.
Email address:* `andrey.ziyatdinov@upc.edu` (Andrey Ziyatdinov)

suitable in early detection scenarios. The full data set is made publicly available to the community.

---

1. Introduction

Animals can actively control different parameters of the respiration cycle such as frequency, amplitude and duration; and thus, adapt the interaction between the stimuli (odourants) and the olfactory receptor neurons (ORN) according to their purposes. For example, respiratory rates of rodents and other small mammals are in the range of $1 - 4 \ Hz$ when rest awake in a safety environment. This frequency increases rapidly (in only one respiration cycle) to $12 \ Hz$ when animals perform odour source localization, novel odourants exploration, or odour discrimination tasks [1], [2].

The respiratory frequency is, therefore, an important feature of the olfactory system when animals sample odours actively since odour perception depends on both the stimulus identity and the sampling method. It has been proposed that rapid sniffing allows the animal to attenuate the odour background and thus facilitate the detection of new odourants [1]. However, how the representation of the olfactory information at the output of the olfactory bulb (OB) changes with the sniff frequency and the computational advantages of high frequency sampling have not been yet elucidated [3].

The active behaviour of animals to extract relevant information from unknown and variable environments inspired active-sensing strategies in the field of artificial chemical sensing. Such strategies have been less explored in chemo-sensory systems than in robotics or computer vision [4]. In an adaptive chemical sensing system, its parameters can be adjusted or tuned in response to the environment, the internal changes and/or the target analytes [4, 5]. In particular the active strategies proposed in chemical systems have been based on either sensor (olfactory receptor) tuning or sampling (air intake) modification. An early example of active sensing system using a gas blender was proposed to determine the composition of gas mixtures from its individual compounds [6]. The component-blending ratio was iteratively adjusted to reproduce the sensor array response to the unknown mixture and resolve the mixture balance.

Metal-oxide (MOX) gas sensors represent one of the most remarkable and well-known examples of tunable sensors. The operation temperature of MOX sensors significantly affects the sensors's selectivity and sensitivity profiles due to changes in the gas reactions that occur at the sensing layer [7, 8]. One can modify the operation temperature to accomplish a certain sensing scenario, for example, to enhance the detection of target analytes in complex mixtures [9, 10, 11]. More recently, the active-sensing strategy was implemented in MOX sensors to reduce the power consumption and minimize the sensing costs in real time via optimization of the operating temperature [12], [13].

Robots equipped with chemical sensors are good examples of chemical sensing applications, where adaptive and biologically inspired strategies work. An odour source localization system equipped with pumps was proposed to actively generate air currents, enhance the difference between responses of right and left gas sensors and, thus, increase system performance [14]. A similar system employed an air curtain also to stimulate the difference between responses of the sensors in accordance with the direction toward a gas source [15]. A crayfish robot with fanning arms and electrochemical sensors was shown to enhance chemical signal reception by actively generating water currents using fanning arms [16]. Results of comparison between biologically plausible models and statistical strategies for odour localization underline the limitations in the sensor technologies and robotic platforms currently available, giving one of the suggestions to take inspiration from the successful principles reflected in

animal behaviour [17, 18].

The problem of sensor-array optimization by signal modulation was addressed by another group of authors [19, 20, 21, 5], where the flow of the carrier gas was modulated by means of PC-controlled pumps. The authors observed new reproducible dynamic patterns in the sensor response and showed an enhancement in sensor selectivity under the classification task. Another work addressed the problem of the long-term drift, and the proposed sampling protocol in the set up by short-time repeated exposures resulted in an improvement of the reproducibility for an array of seven quartz microbalance gas sensors [22]. It is worth noting that an early-detection scenario under the condition of the flow modulation was not explored in that works.

Extraction of informative features available early in the sensor transient signals was a permanent research topic in machine olfaction [23, 24, 25, 26]. It was basically shown an acceleration by using early features, where the conventional sampling method is an exposition to a gas pulse. As far as we can ascertain, conventional transient signals and sensor signals modulated by the gas flow have not been compared yet in terms of early availability of informative features.

In this work we propose a chemical sensing system based on an array of 16 MOX sensors and featured with an external mechanical ventilator to simulate the biological respiration cycle. Our goal is to characterize and explore the sensor signals in response to the modulated gas flow at a fixed respiration frequency. We expect to confirm a superior performance of the proposed system on the early detection scenario. The acquired modulated signals are decomposed into low-frequency and high-frequency components, and the resulted feature sets are compared in terms of the discrimination performance. We assume that the low-frequency part of the modulated signals approximate the signals that would be measured under the standard constant flow rate mode. Such assumption was empirically confirmed in terms of the transient dynamics, which were similar for the filtered low-frequency part of the modulated signal and the signal acquired when the flow modulation was disabled.

Hence, we state two hypotheses. Hypothesis H_1: Does any information valuable for discrimination of the exposed gases exist in the modulated response (comparison between low-frequency and high-frequency components of the modulated response)? We formulate a second hypothesis H_2, conditioned to the H_1: Does this information appear at the early stage of the course of measurement, that in turn can be used for the early detection scenarios in machine olfaction? If both hypotheses are proven, the proposed sensing system featured by gas flow modulation would suit to be a dynamic system targeted to the early detection scenario, having the respiration frequency as a tuning parameter.

2. Materials and Methods

2.1. Experimental Set Up

The developed experimental set up to explore chemical sensing while simulating a biological respiration cycle was composed of three main blocks: an array of commercially available metal-oxide (MOX) sensors with a custom signal conditioning board, an embedded computer for sensor control and signal acquisition, and a custom gas delivery system with a mechanical ventilator for inducing the respiration behaviour in the gas flow. A scheme of the measurement set up is shown in Figure 1.

The array was composed of 16 MOX gas sensors of 5 different TGS models from Figaro Inc. [27]. The sensor devices were mounted on a custom printed circuit board with ground, power supply, heater voltage input and sensor voltage output lines. In order to increase the sensor variability in the array, the sensors were set at two different operating temperatures, which were controlled with the applied voltage (3.3 V or 5 V) on the heating element integrated in the TGS sensors. The sensors' signals were acquired by means of a voltage-divider scheme with a load resistor (21 $K\Omega$ or 82 $K\Omega$). Hence, the setup resulted in 10 different sensor conditioning profiles based on the combination of 5 TGS model, 2 sensor operating temperatures and some sensor redundancy. Table 1 gives the details on the configuration of the sensors. Sensors were selected heuristically so that the sensor transients were able to follow the flow dynamics. The circuit board with the gas sensor array was placed in a 70 ml inner volume chamber connected to the mechanical ventilator.

The acquisition of sensor signals was performed by a PC104-standard embedded computer, that was designed for real-time acquisition, processing and visualization of sensory data for an autonomous mobile robot [28, 29]. The voltage output lines of the circuit board were read by the 16-bit Analog-to-Digital Converter (ADC) board with 16 channels acquiring at a frequency of 25 Hz. The embedded computer ran a custom built GNU/Linux image with the aid of a software component for a refined control of the measurement process.

The central part of the gas flow system was a mechanical ventilator made commercially available from Harvard Apparatus [30]. This device is principally targeted to ventilate animals, ranging from mice to cats, with a precise volume and pressure control of the respiration profiles. The mechanical ventilator includes a cylinder of volume 63.44 cm^3 and a mechanical pump that takes air from the outlet *Source* and pushes the air sample through the outlet *To Animal*. Moreover, the system receives the sample again in the outlet *From Animal* to close the loop, control the air pressure decay, and collect the exhaled air. By means of the mentioned mechanical ventilator, we emulated the respiration behaviour in animals (see Figure 1). The *Source* channel was connected to the vessel with a dilution prepared in advance. Finally, the gas chamber was connected to the channels *To Animal* and *From Animal* to expose the sensors to the prepared sample and collect the sample before being exhausted. The cylinder of the ventilator was fixed to a frequency of 5 breaths per minute, approximately equivalent to 0.08 Hz for all the measurements. The respiration frequency is a critical parameter of the experimental set up, because the frequency has to be low enough for the temporal dynamics of the sensor responses to follow the flow. Figure 2 (left panel) shows that the responses of the sensors follow the respiration cycle and contain the modulated part of the signal.

2.2. Data Set and Measurement Protocol

We tested the approach based on the biologically inspired gas sampling on a mixture quantification problem of two analytes, ethanol and acetone. Three concentrations doses 0.1, 0.3 and 1 vol.% were used to prepare the dilutions in water for the pure analytes and a set of binary mixtures. The gas classes included samples of pure ethanol (labels *eth-0.1*, *eth-0.3* and *eth-1*), samples of pure acetone (*ace-0.1*, *ace-0.3* and *ace-1*), samples of binary mixtures of ethanol and acetone (*ace-0.1-eth-0.1*, *ace-0.1-eth-0.3*, *ace-0.3-eth-0.1*, *ace-0.1-eth-1* and *ace-1-eth-0.1*) and samples of water dilutions without any analyte (*air*) giving a total number of 12 classes. The choice of these analytes and concentrations was not affected by any particular application constraint, except that the sensors of selected models show consistent and diverse responses among the gas classes.

5

We conducted the measurements by splitting them into 6 batches, where each batch contained records approximately for all gas classes given in a random order. That resulted in the total number of records equal to 58. All the batches were collected in a time period of 4 days to minimise the effect of the long-term internal and environmental noise in the system.

Each of the measurements was performed according to the following protocol: using a micropipette we delivered 10 μl of the corresponding dilution to the vessel, which in turn was connected to the apparatus *Source* channel to expose the sensor array to the generated gas sample. After 3 min of exposition, the source of the gas vapour was removed from the vessel to start the recovery phase. During the recovery phase, the system was sampling room air for 2 additional minutes to record the decay in the sensors' signals. The recorded transient signals from the sensors were acquired at the sampling frequency of 25 Hz during 5 min, resulting in 7500 data points for each measurement per sensor. It is important to remark that 2 minutes of recovery phase was not sufficient to recover the sensors' baseline and re-establish again the initial conditions in the gas chamber. Hence, although we acquired 2 minutes of recovery phase, the system was pumping air until the sensors recovered the baseline and the whole gas sample was exhausted from the gas chamber.

The complete acquired data set can be represented as a 3-dimensional array of size 58 $\times$ 7500 $\times$ 16. These dimensions correspond to 58 different measurements, each of them including 7500 data points of the acquired time series for the 16 sensors. The readout data was the output voltage of the sensor stored as resistance values according to the voltage-divider scheme and using the value of the load resistor. Hence, each data point in the array described the resistance of a sensor $R(t)$ at a certain time of measurement t. The resistance values in the data set were normalized by subtracting the baseline value $R_0 = R(t_0)$ at the starting point of the measurement t_0 and scaling by factor R_0, $(R(t) - R_0)/R_0$.

2.3. Feature Sets and Methods

We extracted two different sets of features from the acquired signals to investigate the amount of information that can be extracted from each set of features. In particular, we divided the sensors' signals (which are modulated by the periodic respiration oscillations at $0.08\ Hz$) in low and high frequency components. The first set describes the low-frequency component in the signals, that corresponds to a conventional monotonically increasing/decreasing temporal response to a gas pulse. The second set contains the high-frequency component in the signals related to the respiration stimuli with a oscillatory shape.

The low/high frequency signals were obtained by means of a set of digital filters. First, the acquired signals were pre-processed with a median filter to remove the spikes in the signals. Second, we employed two Butterworth filters of 3rd order: a low-pass filter (cut-off frequency 0.01 Hz) and a high-pass filter (pass-frequency 0.07 Hz) to generate the low/high frequency signals respectively.

For feature extraction we divided both low-frequency and high-frequency sensor signals in respiratory cycles, where each cycle was processed independently. Thus, we will hereinafter refer to feature as *feature by respiratory cycle*. Since high-frequency trajectories showed oscillatory behavior similar to a sine wave curve, we decided to follow a very straightforward strategy for feature extraction. We used amplitude of the high-frequency signal (oscillation) at every respiratory cycle as a feature. Low-frequency trajectories had a monotonic behaviour, and we used the magnitude of the low-frequency signal as a feature at every respiratory cycle. The magnitude value was taken at the same time of oscillation, where the amplitude of the

6

high-frequency signal was measured. We also introduced a cycle-independent feature per single measurement, defined as the maximum of the low-frequency signal over the course of the measurement.

Data analyses were computed on the following data sets (or sub-sets) of features in combination with the information on the analyte concentrations.

- X_{hf}: a 3-dimensional matrix with 58 observations $\times$ 13 high-frequency features $\times$ 16 sensors.

- X_{lf}: a 3-dimensional matrix with 58 observations $\times$ 13 low-frequency features $\times$ 16 sensors.

- X_{max}: a 2-dimensional matrix with 58 observations in rows; 16 maximum features in columns. These cycle-independent features are included in the analysis as a reference to X_{lf} and X_{hf}.

- Y: a 2-dimension matrix with 58 observations in rows; 2 concentration values of acetone and ethanol in columns.

- Y_{lab}: a vector of length 58 that encodes 12 gas class labels derived from Y matrix: *ace-0.1, ace-0.3, ace-1, eth-0.1, eth-0.3, eth-1, ace-0.1-eth-0.1, ace-0.1-eth-0.3, ace-0.3-eth-0.1, ace-0.1-eth-1* and *ace-1-eth-0.1*.

Qualitative comparison of the two sets of features was performed by means of two approaches, principal component analysis and mutual information theoretic measures. Principal component analysis (PCA) is a projection method in multivariate statistics that represents data in low-dimensional space and provides a sequence of best linear approximations to that data [31]. We computed PCA with *pls* R package [32]. In this work we used PCA for two purposes: to visualise trajectories for different gas classes, and to compute a lower-dimensional representation of the complete sets of features. PCA is one of the most popular methods in machine olfaction for representing a data set acquired from an array of gas sensors. One can typically observe that the first two principal components capture the most of the data variance and there are certain variance directions in that multivariate space: direction of the concentration gradient, direction responsible for analyte discrimination and direction containing the drift-related noise.

The Mutual Information is a measure of the amount of information of one random variable contained into another random variable and has already been proposed for optimization or estimation of the limit of detection in chemical systems [33, 34, 35]. The mutual information (MI) between the feature X and the label Y is $MI(Y, X) = H(X) + H(Y) - H(Y; X)$, where $H(X)$ and $H(Y)$ stand for the marginal entropy of X and Y respectively, and $H(Y; X)$ stands for the joint entropy of two variables X and Y. Here, we computed the MI between the random variable that represents the gas label and the random variable of the extracted features (one feature by respiration cycle, of the single sensor). The entropy was calculated as the Shanon entropy formulation in units of *bits*. We used a kernel density estimation for bandwidth selection described in [36] and implemented in *KernSmooth* R package [37]. We examined three probability estimators needed for calculation of the entropy: the empirical estimator, the Miller-Madow estimator [38] and the Schurmann-Grassberger estimator [39],

and selected the most conservative Miller-Madow method. We used the routine on computation of the information measures implemented in *infotheo* R package [40]. Discretization of continuous data was performed by means of the equal width binning algorithm [41].

From the quantitative modeling point of view, the data analysis meant a multivariate regression with multiple responses, where the predictors were the features extracted from the sensor signals and the responses were the concentrations of two analytes, acetone and ethanol. To this purpose we employed partial least squares (PLS) regression method, being one of the standard tools in chemometics [42] and machine olfaction [43, 44]. In particular, we chose the canonical power partial least squares (CPPLS) approach for gaining computational efficiency, among others [45, 46]. All CPPLS models were fitted with $\gamma = 0.75$.

We followed the standard sampling procedure to validate the regression models. We randomly divided the complete dataset composed of 58 measurements into training and testing subsets, with ratios of 80% and 20%, respectively. The number of latent variables for the PLS models was selected in the range $[1, 10]$ through leave-one-out cross-validation over the training samples. We applied the resulted models with the optimized parameters to the test samples and estimated the model performance. Performance for all models is reported through the root-mean-square error in prediction (RMSEP). This process was repeated with random split assignments for 100 times.

In order to quantify the prediction power across time (across the transient), we considered the features extracted since the beginning of the measurement until the time t that corresponds to the k-*th* flow oscillation. In other words, we extracted the features since the injection of the odourants to the k-*th* cycle to quantify the detection ability in time. That results in the total number of $16 \times k$ features (acquired from the signals from 16 sensors at the time t) passed to a quantitative model at the k-*th* cycle. On the contrary, the fixed number of 16 features (one feature per sensor) is considered for k-*th* cycle in the qualitative analysis described above.

3. Results

Figure 2 shows the results of the low-pass and high-pass filtering operation on an example of the measurement of ethanol at 1 vol.% recorded from sensor No 7. The signal given on Figure 2 (a) has been decomposed into slowly changing signal on Figure 2 (b) and oscillatory signal on Figure 2 (c). The low-frequency signal showed a typical transient sensor response to the gas pulse (as if the modulation by the gas flow system did not exist). In the exposition phase the signal was monotonically increasing until the time the saturation regime was reached at approximately 2 min. In the next cleaning phase the signal was monotonically decreasing. The high-frequency signal showed an oscillatory shape with the dominant harmonic matching the respiration frequency 0.08 Hz. The amplitude of the sine wave was increasing at the beginning of both the exposition and cleaning phase, but it was decreasing further as the phase was coming to the end. Interestingly, the high-frequency signal reaches its maximum before saturation in the original signal. Hence, one can observe that the high-frequency component in the sensor signal (when the gas flow system is modulated by the mechanical ventilator) shows a faster response, in terms of the *maximum* signal magnitude, to changes of the gas volume occurring as the measurement goes on. This ability for early detection may be a basis for beneficial performance in different pattern recognition tasks, which take place further in the data processing flow.

It is interesting to characterise the oscillatory signals through a space-state representation. We took all the high-frequency transient signals (7500 data points) for two classes, acetone at 1 vol.% and ethanol at 1 vol.%, and represented the data as trajectories in a Principal Components coordinate system. Figure 3 shows a part of the trajectories from the second respiration cycle with a length equal to the respiration period (334 data points). The first principal components capture most of the variance and represent the trajectories mostly as circles, as shown on Figure 3 (a). The radii of these circles correspond to the oscillation amplitudes at the given respiration cycle. The data projected on higher order components is shown on Figure 3 (b) and Figure 3 (c), where it is clearly observed that the trajectories show a different behaviour in shape, for each of the two classes. High-frequency components of the modulated sensor signals developed a complex shape containing clear class information.

3.1. Qualitative Analysis of Low- and High-Frequency Features

We computed PCA for the three matrices of features X_{max}, X_{lf} and X_{hf} (the last two matrices were converted to 2-dimensional matrices by slicing along the first dimension) and plotted the data scores projected onto the first two components. Figure 4 shows the three plots of PCA scores coloured by the gas class labels stored in Y_{lab} vector. All three sets of features X_{max}, X_{lf} and X_{hf} show a very similar distribution of data points among the gas classes. In particular, the points of pure analytes (pure acetone, red labels: *ace-0.1*, *ace-0.3* and *ace-1*; pure ethanol, blue labels: *eth-0.1*, *eth-0.3* and *eth-1*) convey the boundaries, where the rest of points of binary mixtures are placed. For example, the points of the binary mixtures with higher concentration of ethanol (labels *ace-0.1-eth-1* and *ace-0.1-eth-0.3*) are positioned closer to the points of pure ethanol.

Both sets of features X_{max} and X_{lf} were derived from the same low-frequency signals, but the sets differ in the number of features (1 feature per sensor in X_{max} set and 13 features per sensor in X_{lf}). One may consider that X_{max} set contains the most stable and informative features of the low-frequency signals, while X_{lf} set provides much more features, which are expected to contain some additional information with a certain redundancy across the features. A comparison between two score plots on Figure 4 (a) and (b) (computed for sets X_{max} and X_{lf}, respectively) reveals the effect of adding more features into the analysis. In both plots, the first component of the PCA captures most of the variance and points out the direction of the concentration gradient, while the second principal component captures the class-related variance and clearly indicates two directions, towards acetone (bottom) and ethanol (top). The higher order principal components are likely to contain the data variance related to the noise. The variance captured by first component for X_{max} set (92.51%) is greater than for X_{lf} set (83.5%), which means that the features from X_{lf} set jointly tend either to explain more class-related distribution or to reveal more noise. Another observation is concerned about to group of data points of water class (black labels), which are substantially disperse, as shown on Figure 4 (b). That might be attributed to an instability in X_{lf} set, due to the large amount of features and the redundancy among them. We can conclude that the distribution of the scores (projected on the first two principal components) for two sets X_{max} and X_{lf} is roughly the same for all the gas classes, while the scores for X_{lf} set tend to contain more variability.

The principal component analysis for two sets X_{max} and X_{lf} reveals a similar information about the low-frequency part of data. We can compare the set of the high-frequency features X_{hf} directly with X_{lf} set, as these have the same size. The score plots for X_{lf} and X_{hf} sets

are given on Figure 4 (b) and (c), respectively. One can clearly observe that the variance captured by the first two principal components is more evenly distributed for X_{hf} set. The direction of the variance related to the concentration gradient is mostly aligned along the first component for X_{lf} set, while this direction is somehow shared between the first two components for X_{hf} set. In addition, the data points seem to be more separated among the groups of the gas classes for X_{hf} set. We can conclude that the high-frequency features capture higher orders of variability when compared with low-frequency features.

The results on evaluation of two sets of features by the mutual information measure are presented on Figure 5. Plots (a) and (c) on Figure 5 show all 13 features of sensor No 16 stored in sets X_{lf} and X_{hf}, respectively, where each line corresponds to one single measurement of a particular gas class. In both sets the features are ordered in time with each feature number being equal to the number of respiration. Hence, one can observe the change of feature value in time during 13 respiration cycles. It is important to note that a low-frequency feature of the maximum value (per measurement) occurs at the very end of the course of the measurement while a high-frequency feature of the maximum value occurs at an early stage of the measurement (2-4 respiration cycle).

Plots (b) and (d) on Figure 5 present the values of mutual information computed per feature and per sensor for sets X_{lf} and X_{hf}, respectively. In this heatmap image the rows represent the sensor number, the columns represent the respiration cycle number and the colour encodes the mutual information values. The images underline a trend where the features closer to their maximum value demonstrate the larger values of mutual information. The maximum values of high-frequency features also carried the higher information on the class, as shown on Figure 5 (d). It is important to emphasize the earlier time of accessibility of the feature with the higher mutual information in the case of the high-frequency features.

One can also observe on Figures 5 (b) and (d) that the average colour intensity for the heatmap image for X_{lf} set is higher than for X_{hf} set. This means that the low-frequency features outperforms the low-frequency features in the case of one-by-one comparison, but the difference, for example, between two subsets of the best features from two sets X_{lf} and X_{hf} is inconclusive. We did not address this issue in the computation of mutual information, as the estimation of joint distribution of a set of features requires a larger sample size, otherwise, the bias in estimation would be large due to small sample size. Indeed, we will tackle this problem further in this section by performing a quantitative analysis.

The most clear evidence to support the observation that the most discriminant features appear earlier in set X_{hf} than in sets X_{lf} is found when computing the mutual information in a pairwise manner to compare the features from both sets for a single sensor. Figure 6 presents the heatmap image with the feature number from X_{hf} set in the rows, the feature number from X_{lf} set in the columns and the colour encoding the mutual information value. The image shows a clear tendency of the earlier high-frequency features to strongly correlate with the latest low-frequency features. The discrimination capability of the low-frequency features of sensor No 1 monotonically increases and the features 12 and 13 have the highest values, as shown on Figure 5 (b). One can observe on Figure 6 that the high-frequency feature 3 can be a reasonably good substitute for the low-frequency features 12 and 13.

3.2. Quantitative Analysis on Early Detection Scenario

Figure 7 shows the model performance as a function of time (respiration cycles) for two types of features: low-frequency (red colour) and high-frequency (blue colour). Plots (a) and

(b) on Figure 7 correspond to acetone and ethanol analyte respectively. The dashed black line represents the performance of the reference model computed on the set of maximum features X_{max}. Both plots show the abrupt improvement in the performance on time during the first three respirations, that also demonstrates the weakness of the features available at the first and second respiration in terms of the analyte quantification. The sub-sets of features available after the third and fourth respiration can provide a good performance and reach the performance level of the reference model. The performance line of X_{hf} set (blue colour) shows a better dynamics than the line of X_{lf} set (red colour) and has lower error rates at most of the time points.

4. Conclusions and Discussion

In this work we proposed the experimental set up featured by the mechanical ventilator to modulate the flow in the gas delivery system. The modulated signals acquired from the array of 16 metal-oxide sensors were shown to contain two parts separated in the spectra: the low-frequency signals, which demostrate a conventional response curve of a sensor in response to a gas pulse, and the high-frequency signals, which have a clear principal harmonic at the frequency 0.08 Hz. The tested gas classes formed a relatively broad combination of two analytes, acetone and ethanol, in binary mixtures. For the comparative analyses we constructed two sets of features, low-frequency and high-frequency, with 13 features per sensor ordered in time (respiration cycle).

We have shown that the high-frequency features contain a certain amount of information related to the gas classes. The dimensionality reduction of the high-frequency feature set by PCA suggested that these features differ from the conventional low-frequency features in the sense that the first principal component does not fully contain the variance direction of the concentration gradient and the features form a higher-dimensional space. Evaluation of the two sets of features by the mutual information measure showed that both types of features have a similar capability for discrimination among gas classes, while the most discriminative high-frequency features occur earlier in the course of the measurement.

We have conducted the quantitative analysis to confirm that the discriminative information of the high-frequency features appears at the early stage of the measurement. The PLS-based regression model trained on the set of high-frequency features showed a superior performance on the early stage of the measurement in comparison with the same modeling with the set of the low-frequency features. It is worth noting that the strategy in signal- and data-processing applied to the acquired modulated signals was straightforward in this work. The analysis can be improved in a number of ways, on the stages of feature extraction and/or pattern recognition.

Characterization of the sensors in terms of sensitivity and selectivity could give some insights in the difference between two types of features. In the supplemental material we compared these two sensor characteristics for the use 16 metal-oxide sensors by means of an exploratory data analysis. We fed the data from low-frequency and high-frequency sensor responses (records from the first batch of measurements, samples of pure analytes, and the features extracted on the fourth respiration cycle). Two types of features showed similar normalized sensitivities to the two analytes. The selectivity profiles found in the high-frequency features were more balanced between the analytes, while the selectivity to acetone was higher for most of the low-frequency features compared to ethanol. The later observation might be

11

a possible explanation of better performance of the regression model for acetone rather than for ethanol evaluated on the low-frequency features (Figure 7). Interestingly, the observation of a more even distribution of the selectivities in the high-frequency features might be the reason of a higher dimensionality of the data space formed by the high-frequency features commented in the PCA analysis and, consequently, better performance of the correspondent regression models. Analysis of future data sets collected on the proposed set up is suggested to consider characterization of the sensors in terms of sensitivity, selectivity or other performance indices.

The data analysis presented in this work is not based on any physical sensor model under the conditions of the gas flow modulation. The sought phenomenon due to the flow modulation in the set up can be described as a cycled non-equilibrium working mode of a metal oxide sensor. The concentration of analytes at the surface are periodically changed, and the adsorption-desorption reactions at the surface and the consequent diffusion processes are altered. That in turn leads to the development of the specific response patterns read out from the sensor. As far as we can ascertain, there is only one publication that presented the conductance-transient analysis of thick-film tin oxide gas sensors under successive gas-injection steps [19]. We expect that the results reported in this work will motivate further investigation in the direction of the sensor modeling under different gas flow conditions.

In our point of view, the most prominent strategy to enhance the experimental set up is related to the concept of active sensing [4]. In particular, the parameter that might be manipulated at the design level is the respiration frequency. The proposed system would be truly biologically inspired, at the hardware level, in comparison to those reviewed in [4]. We believe that the approach presented here will provide new opportunities for exploring the complex scenarios in the neuromorphic simulations of machine olfaction and will be very valuable for system integration with the state-of-the-art of neuromorphic systems [47, 48]. Hence, in order to continue the joint effort towards more biologically inspired systems that may show better performance than other approaches, we decided to publish the data set for free public use. The complete data set is made available on the web site of The University of California at Irvine (UCI) Machine Learning Repository, while the results reported in this work should be considered as a reference.

Acknowledgments

References

[1] J. V. Verhagen, D. W. Wesson, T. I. Netoff, J. a. White, M. Wachowiak, Sniffing controls an adaptive filter of sensory input to the olfactory bulb., Nature neuroscience 10 (5) (2007) 631–9. doi:10.1038/nn1892.
URL http://www.ncbi.nlm.nih.gov/pubmed/17450136

[2] L. M. Kay, J. Beshel, J. Brea, C. Martin, D. Rojas-Líbano, N. Kopell, Olfactory oscillations: the what, how and what for., Trends in neurosciences 32 (4) (2009) 207–14. doi:10.1016/j.tins.2008.11.008.
URL http://www.ncbi.nlm.nih.gov/pubmed/19243843

[3] M. Wachowiak, Active Sensing in Olfaction., in: Menini A. (Ed.), The Neurobiology of Olfaction., boca raton Edition, CRC Press, 2010.
URL http://www.ncbi.nlm.nih.gov/pubmed/21882431

[4] R. Gutierrez-Osuna, A. Hierlemann, Adaptive microsensor systems, Annual review of analytical chemistry (Palo Alto, Calif.) 3 (February) (2010) 255–276. doi:10.1146/annurev.anchem.111808.073620.
URL http://www.ncbi.nlm.nih.gov/pubmed/20636042

[5] A. Vergara, E. Llobet, Sensor selection and chemo-sensory optimization: toward an adaptable chemo-sensory system., Frontiers in neuroengineering 4 (January) (2011) 19. doi:10.3389/fneng.2011.00019.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3250696&tool=pmcentrez

[6] T. Nakamoto, H. Matsushita, N. Okazaki, Improvement of optimization algorithm in active gas/odor sensing system, Sensors and actuators. A, Physical 50 (3) (1995) 191–196.
URL http://cat.inist.fr/?aModele=afficheN&cpsidt=5974464

[7] A. P. Lee, B. J. Reedy, Temperature modulation in semiconductor gas sensing (1999) 35–42.

[8] F. Wlodek, Stanislaw and Colbow, Konrad and Consadori, Signal-shape cycled tin-oxide gas sensor, Sensors & Actuators: B. Chemical 3 (1991) 63–68.

[9] T. A. Kunt, T. J. Mcavoy, R. E. Cavicchi, S. Semancik, Optimization of temperature programmed sensing for gas identification using micro-hotplate sensors, Sensors & Actuators: B. Chemical 53 (1998) 24–43.

[10] D. Meier, J. Evju, Z. Boger, B. Raman, K. Benkstein, C. Martinez, C. Montgomery, S. Semancik, The potential for and challenges of detecting chemical hazards with temperature-programmed microsensors, Sensors and Actuators B: Chemical 121 (1) (2007) 282–294. doi:10.1016/j.snb.2006.09.050.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400506006551

[11] B. Raman, D. C. Meier, J. K. Evju, S. Semancik, Designing and optimizing microsensor arrays for recognizing chemical hazards in complex environments, Sensors and Actuators B: Chemical 137 (2) (2009) 617–629. doi:10.1016/j.snb.2008.11.053.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400508008125

[12] R. Gosangi, R. Gutierrez-osuna, Energy-Aware Active Chemical Sensing, in: IEEE Sensors Journal, 2010, pp. 1094–1099.

[13] R. Gosangi, R. Gutierrez-osuna, Active Temperature Programming for Metal-Oxide Chemoresistors, IEEE Sensors Journal 10 (6) (2010) 1075–1082.

[14] A. Kohnotoh, H. Ishida, Active Stereo Olfactory Sensing System for Localization of Gas/Odor Source, 2008 Seventh International Conference on Machine Learning and Applications (2008) 476–481doi:10.1109/ICMLA.2008.101.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4725016

[15] I. Miyatani, H. Ishida, Active Stereo Nose : Using Air Curtain to Enhance the Directivity (21360113) (2010) 1522–1525.

[16] R. Takemura, K. Takahashi, T. Makishita, H. Ishida, Adaptive Chemical Sampling Device Inspired by Crayfish, ECS Transactions 50 (12) (2013) 259–266.

[17] H. Ishida, Y. Wada, H. Matsukura, Chemical Sensing in Robotic Applications: A Review, IEEE Sensors Journal 12 (11) (2012) 3163–3173. doi:10.1109/JSEN.2012.2208740.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6247459

[18] V. Hernandez Bennetts, A. J. Lilienthal, P. P. Neumann, M. Trincavelli, Mobile robots for localizing gas emission sources on landfill sites: is bio-inspiration the way to go?, Frontiers in neuroengineering 4 (January) (2011) 20. doi:10.3389/fneng.2011.00020.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3268183&tool=pmcentrez

[19] E. Llobet, X. Vilanova, J. Brezmes, R. Alcubilla, J. Calderer, J. Sueiras, X. Correig, Conductance-transient analysis of thick-film tin oxide gas sensors under successive gas-injection steps, Measurement Science and Technology 8 (10) (1997) 1133.

[20] C. Duran, J. Brezmes, E. Llobet, X. Vilanova, X. Correig, Enhancing sensor selectivity through flow modulation, in: Sensors, 2005 IEEE, 2005. doi:10.1109/ICSENS.2005.1597727.
URL http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1597727&url=http%3A%2F%2Fi

[21] N. El Barbri, C. Duran, J. Brezmes, N. Cañellas, J. L. Ramírez, B. Bouchikhi, E. Llobet, Selectivity Enhancement in Multisensor Systems Using Flow Modulation Techniques, Sensors 8 (11) (2008) 7369–7379. doi:10.3390/s8117369.
URL http://www.mdpi.com/1424-8220/8/11/7369/

[22] E. Martinelli, M. Santonico, G. Pennazza, R. Paolesse, A. DAmico, C. Di Natale, Short time gas delivery pattern improves long-term sensor reproducibility, Sensors and Actuators B: Chemical 156 (2) (2011) 753–759. doi:10.1016/j.snb.2011.02.034.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400511001456

[23] C. Di Natale, S. Marco, F. Davide, A. D'Amico, Sensor-array calibration time reduction by dynamic modelling, Sensors and Actuators B: Chemical 25 (1-3) (1995) 578–583. doi:10.1016/0925-4005(95)85126-7.
URL http://linkinghub.elsevier.com/retrieve/pii/0925400595851267

[24] E. Llobet, J. Brezmes, X. Vilanova, J. E. Sueiras, X. Correig, Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array, Sensors and Actuators B: Chemical 41 (1-3) (1997) 13–21. doi:10.1016/S0925-4005(97)80272-9.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400597802729

[25] M. K. Muezzinoglu, A. Vergara, R. Huerta, N. Rulkov, M. I. Rabinovich, A. Selverston, H. D. Abarbanel, Acceleration of chemo-sensory information processing using transient features, Sensors and Actuators B: Chemical 137 (2) (2009) 507–512. doi:10.1016/j.snb.2008.10.065.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400508007119

[26] S. Vembu, A. Vergara, M. K. Muezzinoglu, R. Huerta, On time series features and kernels for machine olfaction, Sensors and Actuators B: Chemical 174 (2012) 535–546.

[27] J. Figaro Engineering Inc., http://www.figaro.co.jp/.
URL http://www.figaro.co.jp/

[28] P. F. M. J. V. Vicky Vouloutsi, Lucas L. Lopez-Serrano, Zenon Mathews, Alex Escuredo Chimeno, Andrey Ziyatdinov, Alexandre Perera i Lluna, Sergi Bermúdez i Badia, The Synthetic Moth: A Novel Neuromorphic Approach towards Artificial Olfaction in Robots, in: Neuromorphic Olfaction, CRC Press, 2012.
URL http://www.crcpress.com/product/isbn/9781439871713

[29] S. Marco, A. Gutiérrez-Gálvez, A. Lansner, D. Martinez, J. P. Rospars, R. Beccherelli, A. Perera, T. C. Pearce, P. F. M. J. Verschure, K. Persaud, A biomimetic approach to machine olfaction, featuring a very large-scale chemical sensor array and embedded neuro-bio-inspired computation, Microsystem Technologies (2013) 1–14doi:10.1007/s00542-013-2020-8.
URL http://link.springer.com/10.1007/s00542-013-2020-8

[30] Harvard Apparatus, Harvard Inspira Advanced Safety Ventilator Manual, Tech. rep. (2003).

[31] J. F. Trevor Hastie, Robert Tibshirani, The Elements of Statistical Learning: Data Mining, Inference, and Prediction., second edi Edition, Springer, 2009.
URL http://www-stat.stanford.edu/ tibs/ElemStatLearn/

[32] B. r.-H. Mevik, R. Wehrens, K. H. Liland, pls: Partial Least Squares and Principal Component regression, R package version 2.3-0 (2011).
URL http://cran.r-project.org/package=pls

[33] J. Fonollosa, L. Fernández, R. Huerta, A. Gutiérrez-Gálvez, S. Marco, Temperature optimization of metal oxide sensor arrays using Mutual Information, Sensors and Actuators B: Chemicaldoi:10.1016/j.snb.2012.12.026.
URL http://linkinghub.elsevier.com/retrieve/pii/S0925400512013354

[34] X. R. Wang, J. T. Lizier, T. Nowotny, A. Z. Berna, M. Prokopenko, S. C. Trowell, Feature selection for chemical sensor arrays using mutual information., PLoS ONE 9 (3)

(2014) e89840. doi:10.1371/journal.pone.0089840.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942325&tool=pmcentrez

[35] J. Fonollosa, A. Vergara, R. Huerta, S. Marco, Estimation of the limit of detection using information theory measures., Analytica chimica acta 810C (2014) 1–9. doi:10.1016/j.aca.2013.10.030.
URL http://www.ncbi.nlm.nih.gov/pubmed/24439498

[36] M. P. Wand, Data-Based Choice of Histogram Bin Width, The American Statistician 51 (1995) 59–64.

[37] Matt Wand, KernSmooth: Functions for kernel smoothing for Wand & Jones (1995) (2013).
URL http://cran.r-project.org/package=KernSmooth

[38] G. A. Miller, Note on the bias of information estimates, Information theory in psychology: Problems and methods 2 (1955) 95–100.

[39] T. Schurmann, P. Grassberger, Entropy estimation of symbol sequences., Chaos (Woodbury, N.Y.) 6 (3) (1996) 414–427. doi:10.1063/1.166191.
URL http://www.ncbi.nlm.nih.gov/pubmed/12780271

[40] P. E. Meyer, infotheo: Information-Theoretic Measures (2013).
URL http://cran.r-project.org/package=infotheo

[41] P. E. Meyer, Information-Theoretic Variable Selection and Network Inference from Microarray Data, Phd, Universite Libre de Bruxelles (2008).

[42] S. Wold, M. Sjostrom, PLS-regression : a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2) (2001) 109–130.

[43] R. Gutiérrez-Osuna, Pattern analysis for machine olfaction: a review, IEEE Sensors Journal 2 (3) (2002) 189–202. doi:10.1109/JSEN.2002.800688.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1021060

[44] S. Marco, A. Gutiérrez-Gálvez, Signal and Data Processing for Machine Olfaction and Chemical Sensing : A Review, IEEE Sensors Journal 12 (11) (2012) 3189–3214.
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6183455&tag=1

[45] S. de Jong, B. M. Wise, N. L. Ricker, Canonical partial least squares and continuum power regression, Journal of Chemometrics 15 (2) (2001) 85–100.

[46] U. G. Indahl, K. H. Liland, T. Naes, Canonical partial least squares-a unified PLS approach to classification and regression problems, Journal of Chemometrics 23 (9) (2009) 495–504. doi:10.1002/cem.1243.
URL http://doi.wiley.com/10.1002/cem.1243

[47] E. Neftci, J. Binas, U. Rutishauser, E. Chicca, G. Indiveri, R. J. Douglas, Synthesizing cognition in neuromorphic electronic systems, Proceedings of the National Academy of Sciences 110 (37) (2013) E3468—-E3476.

[48] M. Schmuker, T. Pfeil, M. P. Nawrot, A neuromorphic network for generic multivariate data classification, Proceedings of the National Academy of Sciences (2014) 201303053.

Tables

	Sensor label	Model	Voltage, V	Load Resistor, KOms
1	R1	TGS 2610	5.0	21
2	R2	TGS 2610	5.0	21
3	R3	TGS 2602	5.0	21
4	R4	TGS 2600	5.0	21
5	R5	TGS 2610	5.0	21
6	R6	TGS 2611	5.0	21
7	R7	TGS 2610	5.0	21
8	R8	TGS 2620	5.0	21
9	R9	TGS 2610	3.3	82
10	R10	TGS 2620	3.3	82
11	R11	TGS 2602	3.3	82
12	R12	TGS 2611	3.3	82
13	R13	TGS 2610	3.3	82
14	R14	TGS 2610	3.3	82
15	R15	TGS 2610	3.3	82
16	R16	TGS 2600	3.3	82

Table 1: Configuration of 16 metal-oxide sensors used in the experimental set up is presented. The sensors differ in the Figaro model [27], the heater voltage and the load resistor. In total, 10 different configurations are available.
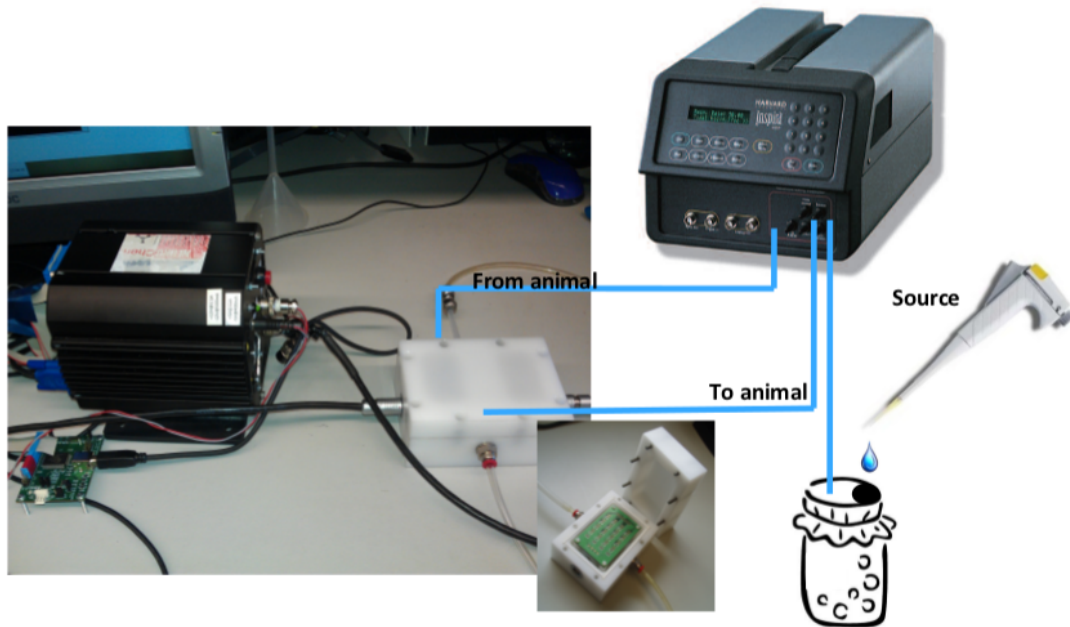
Figures

Figure 1: The experimental set up is composed of three main building blocks: (a) the array of 16 MOX sensors, (b) the embedded computer for controlling the measurement process, and (c) the custom gas delivery system with the mechanical ventilator, which simulates the respiration behaviour in the gas flow. In result, the output signals of the sensors are modulated by the respiration stimuli and contain a high-frequency part in the spectra with a peak placed at 0.08 Hz (the respiration frequency).
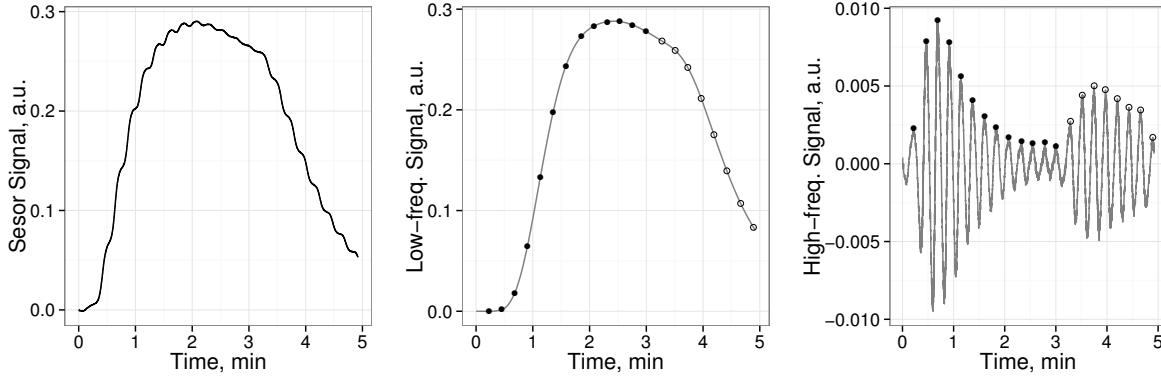
Figure 2: Example of splitting the sensor signal (a) into low-frequency (b) and high-frequency (c) components is shown. The signal was recorded from sensor No 7 in response to ethanol at 1 vol.%. The low-pass Butterworth filter with cut-off frequency 0.01 Hz and the high-pass filter with pass-frequency 0.07 Hz were used. The solid dots on (b) and (c) panels shows 13 features per component, which correspond to the exposition phase of the measurement (length of 2 min) and were used further in the qualitative and quantitative data analyses.
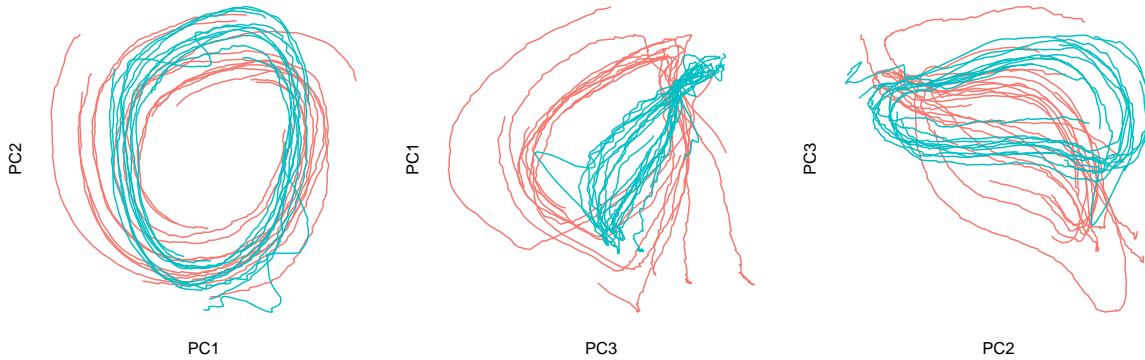
Figure 3: A representation of the high-frequency component of the transient trajectories from 16 sensors is shown in three sub-spaces computed by the principal component analyses (PCA). The data of 334 points (the first respiration cycle) for two gas classes, acetone at 1 vol.% and ethanol at 1 vol.%, were used in calculations. Panel (a) shows that the first two principal components capture the most of the variance and represent the trajectories as circles. A class-specific shapes are presented on panels (b) and (c), where the third principal component is involved.
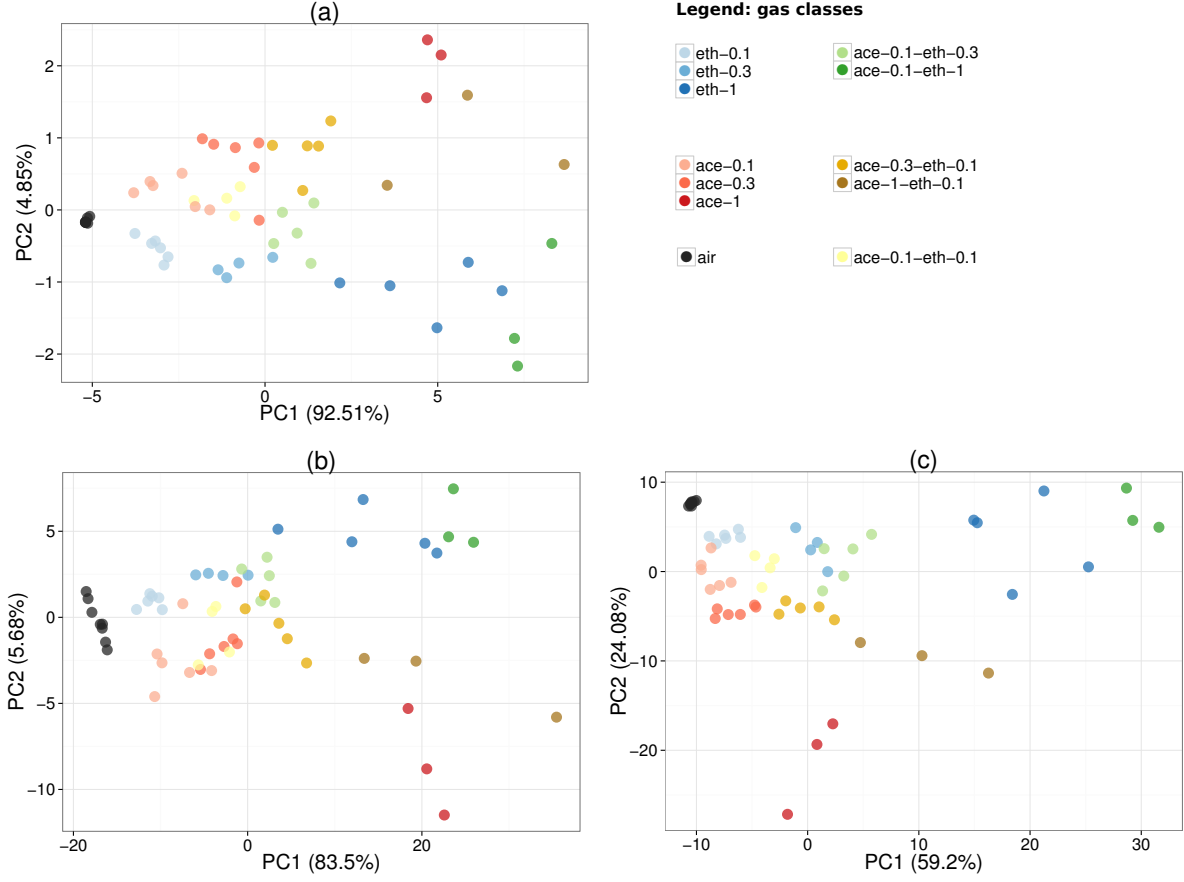
Figure 4: PCA score plots of the principal component analyses (PCA) for three data sets of features computed in this study, (a) X_{max} of the maximum features, (b) X_{lf} of the low-frequency features, and (c) X_{hf} of the high-frequency features, are presented. The data distribution on panels (a) and (b) is very similar in the sense that the first principal component captures the most of the variance and expresses the direction of the concentration gradient. The high-frequency features on panel (c) have slightly different distribution and seem to come from a higher dimensional data space than the low-frequency features on panels (a) and (b), as one can see on the axes labels with the percentage of the captured variance.
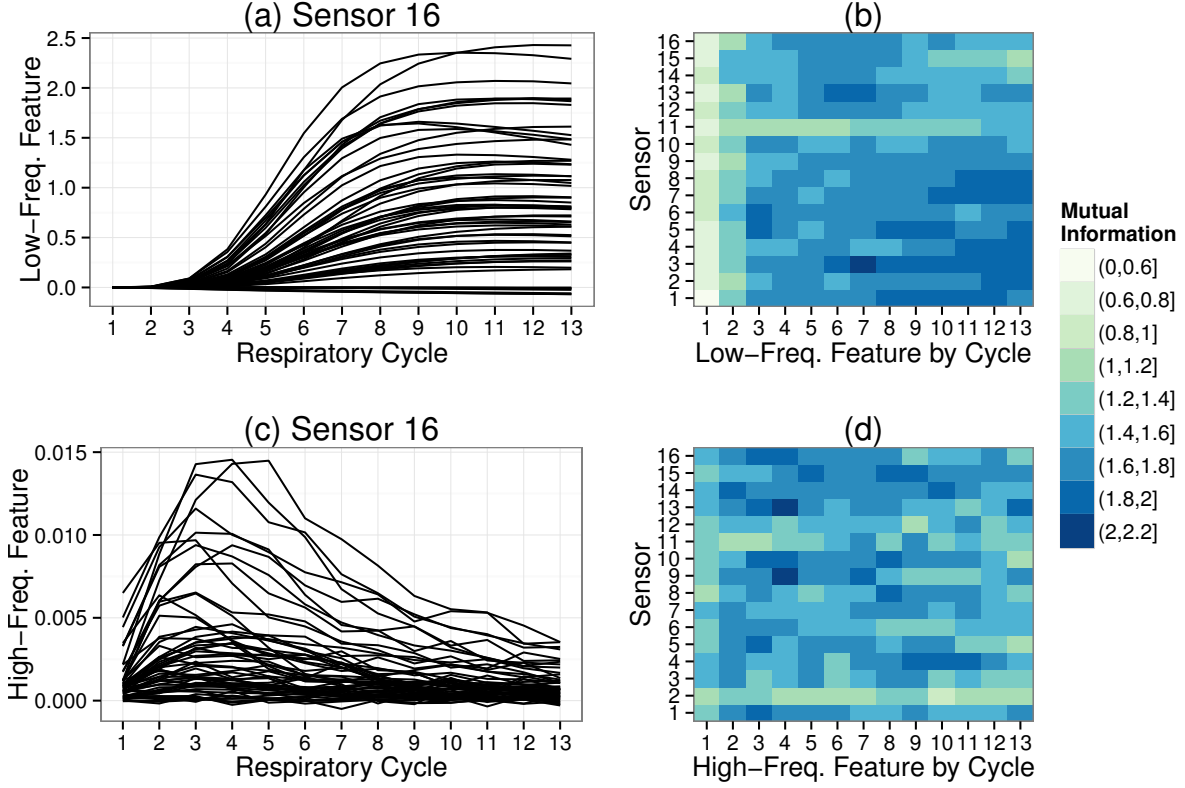
Figure 5: Comparison of two sets of features, X_{lf} of the low-frequency features (upper row) and X_{hf} of the high-frequency features (lower row), in terms of the mutual information measure is presented. The examined 13 features per set are shown on panel (a) for the X_{lf} set and on panel (b) for the X_{hf} set (a single measurement from sensor No 16 in response to ethanol at 1 vol.%). Heatmaps on panels (b) and (d) represent the values of mutual information computed per feature and per sensor for sets X_{lf} and X_{hf}, respectively.
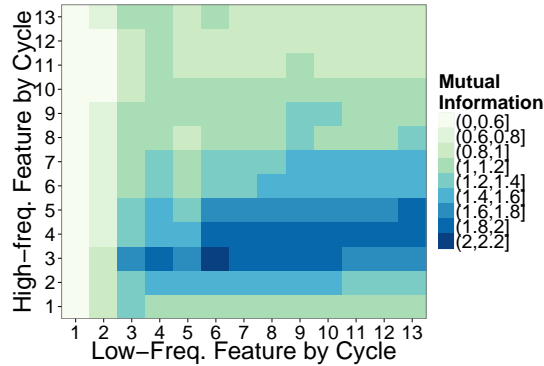

Figure 6: Direct comparison between two sets of features, X_{lf} of the low-frequency features and X_{hf} of the high-frequency features, in terms of the mutual information measure is shown. The image shows that the earlier high-frequency features strongly correlate with the latest low-frequency features.
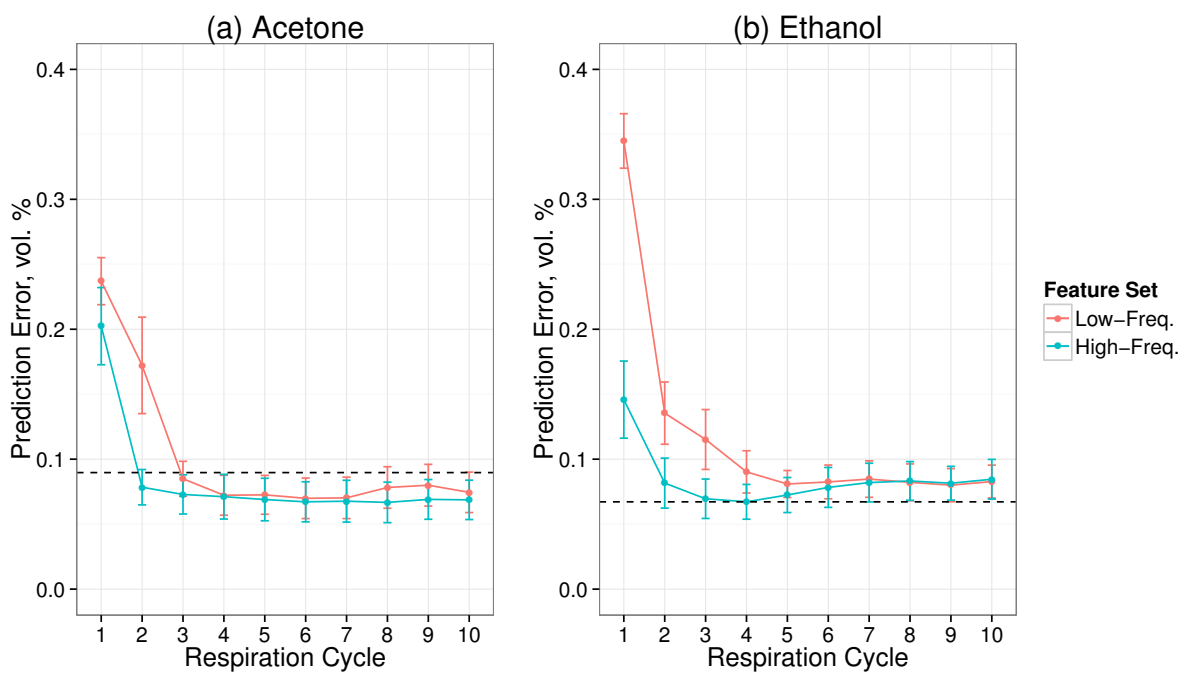
22

Figure 7: The regression model performance as a function of time (number of respirations) for two types of features, low-frequency (red colour) and high-frequency (blue colour), is shown. The error in prediction was measured in vol. % for concentration of two analytes, acetone on panel (a) and ethanol on panel (b). The dashed black line corresponds to the performance of the reference model computed on the set of maximum features.