Odour Classification System for Continuous Monitoring Applications

Marco Trincavelli[a], Silvia Coradeschi[a], Amy Loutfi[a]

[a] *Center for Applied Autonomous Sensor Systems*
Örebro University
Örebro, SE - 70182, Sweden

Abstract

In this paper, we investigate the classification performance of an electronic nose system, based on tin dioxide gas sensors. In contrast to previous studies, the electronic nose is mounted on a mobile platform and samples are analyzed using only transient information in the signals. The motivation behind this work is to explore the feasibility of using electronic nose devices for odour classification in a number of future application domains which require fast and possibly real-time odour identification. To perform transient based analysis of the signals, a comparative study of different methods for feature extraction was performed. Additionally, the application of a relevance vector machine classifier is explored to further analyze the classification performance based on quality of the obtained samples. The results presented in this study can be used for the development of electronic nose devices particularly suitable for environmental monitoring applications.

Key words: chemical sensors array, odour classification, mobile olfaction, relevance vector machines

1. INTRODUCTION

In the past decade, odour discrimination with electronic noses has received growing attention and many studies have been done on how to classify odours using an array of gas sensors and a pattern recognition algorithm.

Email addresses: marco.trincavelli@aass.oru.se (Marco Trincavelli), silvia.coradeschi@aass.oru.se (Silvia Coradeschi), amy.loutfi@aass.oru.se (Amy Loutfi)

Electronic noses are useful for a number of applications in food and perfume industries for quality control or detection of deteriorated goods (1). Normally the sampling procedure occurs in three phases where the sensor array is first exposed to a reference gas, then to the analyte, and finally the analyte is flushed from the system allowing the sensors to recover. Although the three-phase sampling technique has shown to be successfull for a subset of applications, other applications such as online monitoring of industrial processes, pollution monitoring, exploration of hazardous areas, demining or search and rescue may require to sample the environment continuously and/or over longer period of time. These applications may use a sensor network or a mobile robot to collect signals and may involve tasks involving odour-based navigation, gas distribution mapping or gas source declaration (2; 3; 4; 5).

There are many differences between a signal collected with three-phase sampling, in which the sensors are in a chamber where humidity, temperature and exposure to the analyte are controlled, and one collected with an e-nose that samples the environment continuously, where the sensed environment is uncontrolled and the odour is transported according to complex mechanisms that are not fully understood. The most prominent of these differences is that in a real-world and dynamic environment, the steady state of the sensors is almost never reached and consequently the analysis is based on the transient phase of the signal.

Although in the current literature most of the proposed systems use features based on the steady state of the signal, some work has been devoted to analyze how classification can be improved by adding features extracted from the transient phase. In particular, it has been shown in (6) that the features based on both the transient phase and the steady state of the signal outperform the ones based only on the steady state value. Different approaches have been proposed for modelling transient responses. In (7) the transient response has been modeled as a multi-exponential function, while in (6) features extracted fitting polynomial and exponential functions, ARX models, and applying simple heuristics are compared. More recently Carmel et al. (8) proposed a method for modeling the signal directly derived from the physical description of the measurement system. In (9) a feature extraction method based on wavelet transform is described and evaluated. In all these works transient analysis is performed together with steady state analysis and it is not clear how classification performance would be affected when based solely on the transient.

To our knowledge little attention has been given to the problem of clas-

sification when continuously sampling the environment. Martinez et al. (10) proposed a biologically inspired approach for a system able to navigate to a specific gas source in a turbulent environment. The classification algorithm is based on spiking neural networks and uses the steady state information from the gas sensors which are collected when the robot is stopped. In (11) the problem of gas distribution mapping in presence of multiple gas sources emitting different odours was considered. Similarly to the work considered here only the transients were used for classification, however the main focus in these works has been geared towards the navigation or the accuracy of the mapping more than the classification performance per se.

In this paper, we take inspiration from the techniques used to analyze the transient response of signals collected with three-phase sampling technique and examine the application of a new algorithm, namely the Relevance Vector Machine, when using a continuous sampling method. Furthermore, as the sensor array is directly exposed to the environment during the sampling process, a contribution of this work also includes an in-depth analysis of the classification performance and feature extraction techniques suited for this new sampling technique. The paper is organized as follows: in Section 2 we present the classification algorithm. In Section 3 we describe the experimental setup that we used to validate our approach and in Section 4 we present the results of the experiments. Finally in Section 5 we draw some conclusions and give an outline of the future works.

2. CLASSIFICATION OF ODORS USING TRANSIENT RESPONSE

To illustrate the difference between signals collected with three-phase sampling versus continuous sampling an example is shown in Fig. 1. In the upper part of Fig. 1, the time of exposure of the sensors to the analyte is controlled and is long enough to allow the sensors to reach the steady state. Therefore the three phases of the sampling procedure, namely rise, steady state and decay, are well defined and distinguishable. Instead in the lower part of Fig. 1, the signal is collected with the same electronic nose mounted on a mobile robot. The robot is moving at a constant speed and continuously samples the environment. Note that steady state is never reached because the sensors are not continuously exposed for a long enough time to the analyte. This is mainly due to the turbulent nature of the airflow that prevents the gas to distribute uniformly and creates instead a complex structure made of patches and meanders (12).

As a result the main innovation step that is needed with respect to a traditional odour classification algorithm is the introduction of the segmentation step that is able to identify the different phases of the signal. The proposed classification algorithm is articulated in a total of five phases, namely baseline manipulation, segmentation, feature extraction, data normalization and classification. The next subsections describe each of the phases of the algorithm.

2.1. Baseline Manipulation

In order to suppress the noise due to sampling and quantization the signals are smoothed with an averaging filter of dimension five when a sampling rate of 1.25 Hz is used. Then a baseline manipulation is performed in order to minimize the effect of temperature, humidity and short-term drift (13; 14).

The baseline is given as a sequence of readings where there is no significant change in the sensors response. For example in Fig. 1 the baseline is the first initial readings where the nose is stationary and is exposed only to clean air. The most diffused techniques for baseline correction are differential (1), relative (2) and fractional (3):

$$X_{n,s} = R_{n,s} - R_{o,s} \tag{1}$$

$$X_{n,s} = R_{n,s}/R_{o,s} \tag{2}$$

$$X_{n,s} = \frac{R_{n,s} - R_{o,s}}{R_{o,s}} \tag{3}$$

where $R_{n,s}$ is the n^{th} reading for sensor s and $R_{o,s}$ is the mean value of the baseline response. For metal-oxide gas sensors the relative technique is the most commonly used since their resistance at steady state is related to the baseline value through a multiplicative factor (13). When continuously sampling, especially with an open sensing array exposed directly to the environment, the sensors seldomly reach the steady state, and therefore all the three approaches will be evaluated with respect to their effect on the classification performance.

2.2. Segmentation

In order to segment the dynamic response from the signals collected from the nose, we consider the first derivative of the smoothed signal ds/dt, for each gas sensor, s, and characterize three phases of the sensor response as:

- Baseline when $-\epsilon < \frac{ds}{dt} < \epsilon$
- Rising phase when $\frac{ds}{dt} > \epsilon$
- Decay phase when $\frac{ds}{dt} < -\epsilon$

where ϵ is a heuristic threshold to discard smaller changes in the signal due to random oscillation in the sensor response. According to (7) information relevant for gas discrimination can be extracted both from the steady state and the dynamic response of the sensor. However, since it can not be assumed that the sensors reach a steady state phase, only the dynamic response is considered both in the rise and decay phase in order to investigate which of them contains most of the information with respect to classification. Each segmented response is represented by:

$$K_{t,s} = [X_{n,s} \dots X_{n+m,s}] \tag{4}$$

for $t = 1 \dots T$ responses with $X_{n,s}$ and $X_{n+m,s}$ representing the first and the last value of the segmented response.

After segmentation, the possibility of transients containing too little information for achieving reliable classification should be considered. This is done in the classification step by using a classifier that can provide a confidence measure for the classification result. Further information on the classification is described in Section 2.5.

2.3. Feature Extraction

As previously discussed in (15; 9; 6) classification using transient responses has mainly focused on three different feature extraction techniques namely, curve fitting(CF), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) (16; 17).

The procedure of fitting a curve to a time series consists in finding the values of the curve parameters for which the curve approximates best the series of samples. To evaluate which family of curves provides the best performance in fitting the data we use the Root Mean Squared Error (RMSE) that is defined by equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x_i})^2}{n}} \tag{5}$$

where x_i are the measured data and $\hat{x}_i$ are the points estimated by the curve and n is the number of samples we consider in the fitting. The points x_i used for the fitting are the one isolated by the segmentation algorithm. The optimal values of the parameters are then collected in an array that is the feature vector used in the consequent stages of the algorithm. The number and nature of the coefficients depend on the family of curves that are considered for the fitting.

Typically for three-phase sampling systems the sum of exponentials is used since the sensors array is exposed in a stepwise fashion to the odour (15). In this case a first-oder system is appropriate to model the sensor response. In the case of continuous sampling the sensor array is exposed to patches of gas with varying concentrations for varying periods of time. As a result a first-order sensor model is not suitable and a more generic polynomial approximation of the form:

$$Y = \sum_{i=0}^{N} A_i x^i \tag{6}$$

is used. In particular, N is the degree of the polynomial that is considered for the fitting which will be determined experimentally. As it is shown in Section 4, the choice of N is a tradeoff between the accuracy of the fit and the simplicity of the model.

In addition to curve fitting (CF), we also consider the Discrete Fourier Transform (DFT) and the Discrete Wavelet Transform (DWT). The Discrete Fourier Transform provides a description of the signal in the frequency domain, under the assumption that the properties of the signal do not vary during the time interval (stationary signal). The Discrete Wavelet transform instead produces a description of the signal in the time-scale domain and is therefore able to capture abrupt changes in the dynamics of the signal. The DWT is a multilevel decomposition technique that describes a signal in terms of approximation and detail coefficients calculated for different scales of a generating function called a mother wavelet. The DWT can be seen as a generalization of the DFT since the mother wavelet, in contrapposition to the sine function, is not a periodical function and therefore preserves also the location in time of the frequencies it models. Therefore the DWT provides a representation that is richer but less compact than the DFT. Since it is not clear which of the two methods is best suited to the response from the gas sensors, both DWT and DFT are computed for each segmented transient

and fed into the classification algorithm. A comparison of the feature extraction method with respect to the classification performance is given in Section 4. For the DFT the modulus of the first fifteen coefficients are considered. For the DWT the approximation coefficients for the first level decomposition performed with the Daubechies family wavelet are used.

We collect the CF, DFT or DWT for each transient response in a feature vector represented as:

$$F_i = [D_{i,1} \dots D_{i,S}] \tag{7}$$

for $1 \dots I$ transients with $D_{i,s}$ representing either the DWT, the DFT or the best fit coefficients for sensor s in transient i.

2.4. Data Normalization

The data normalization step is aimed at smoothing the sample to sample variations, providing a more regular input to the classification module. Three different kinds of transformations described by the following equations have been applied to this purpose:

$$\phi_n(m) = \frac{F_n(m)}{\sqrt{\sum_{n=1}^{N} F_n(m)^2}} \qquad \forall m, n \tag{8}$$

$$\phi_n(m) = \frac{F_n(m) - \mu_n}{\sigma_n} \qquad \forall m, n \tag{9}$$

$$with \qquad \mu_n = \frac{1}{M} \sum_{m=1}^{M} F_n(m)$$

$$and \qquad \sigma_n = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (F_n(m) - \mu_n)^2}$$

$$\phi_n(m) = \frac{F_n(m) - \mu_m}{\sigma_m} \qquad \forall m, n \tag{10}$$

$$with \qquad \mu_m = \frac{1}{N} \sum_{n=1}^{N} F_n(m)$$

$$and \qquad \sigma_m = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (F_n(m) - \mu_m)^2}$$

where $F_n(m)$ is the value of feature m for sample n, $\phi_n(m)$ is the normalized value of feature m for sample n, N is the total number of samples and M is the dimensionality of the feature vector. Equation 8 is called Vector Normalization since it normalizes the length of every vector to a unitary length. Equation 9 and Equation 10 are referred respectively as Dimension Autoscaling (DAS) and Vector Autoscaling (VAS). In the case of Dimension Autoscaling every *feature* has zero mean and unit variance over its dimensions, where as for Vector Autoscaling every *sample* has zero mean and unit variance over its dimensions.

2.5. Classification

Two main classification approaches have been proposed in literature, one based on statistical classifiers that make use of machine learning and pattern recognition techniques like discriminant functions and density estimation models, and one biologically inspired, that tries to emulate the olfactory system of animals (18). Among these techniques the most applied are the

statistical based classifiers especially neural networks and density models like KNN or gaussian mixture models. In the last decade another statistically based classifier has received growing attention, the Support Vector Machine (SVM). Its strong points are the quick training and the fact that the function minimized during the training process is convex and therefore any local minimum that is found is also global.

The SVM represents the decision boundary in terms of a subset of the training patterns called support vectors. The boundary is selected in order to maximize the margin, that is the distance between the decision boundary and the closest data points. Many variants of SVM have been developed both for classification and regression (19), and the one used here is the soft margin SVM with a radial basis function kernel that was introduced by Vapnik (20) and is an extension of the basic SVM for problems that are not linearly separable. This model has two parameters, the variance of the gaussian kernel and the penalty coefficient for the misclassified samples. The SVM has originally been developed for binary classification but many multiclass extensions have been proposed. According to the results obtained in (21) the approach selected is one-against-one multiclass.

One of the main drawbacks of the SVM is that as a discriminant model it does not provide any posterior probabilities estimation. The estimation of the posterior probabilities is important because it can be considered as a confidence measure of the decision made by the classifier. This opens the possibility to introduce a rejection class, to which all the samples with no definitive membership are assigned. The following equation expresses this concept formally:

$$L_x = \begin{cases} \underset{k}{argmax} P(C_k|x) & \text{if} \quad P(C_k|x) \geq \Gamma \\ rejected & \text{if} \quad P(C_k|x) < \Gamma \end{cases} \tag{11}$$

where L_x is the output label of the classifier for input vector x and $P(C_k|x)$ is the posterior probability of class k given input vector x. Γ is the rejection threshold.

The possibility to reject some of the samples is especially important with respect to a possible suboptimal segmentation policy which may cause that some of the transients do not contain enough information to be classified reliably. A confidence measure for classification can be obtained considering another sparse kernel machine, namely the Relevance Vector Machine.

The RVM has been introduced by Tipping in (22) and has a number of advantages over the more commonly used SVM. Apart from the estimation of the posterior probabilities, the RVM produces a model that is more sparse and therefore quicker in the prediction phase. Furthermore, in the RVM the extension to the multi-class case is more principled than for the SVM (22). This is obtained by coupling the linear predictors y_m for every class m with a softmax function of the form:

$$p(C_m|x) = \frac{e^{y_m}}{\sum_j e^{y_j}} \qquad with \tag{12}$$

$$y_m = \sum_{n=1}^{N} w_{mn} K(x, x_n) + b_m \tag{13}$$

Here $p(C_m|x)$ is the posterior probability of class m given the feature vector x, N is the number of training samples, K is the kernel function, w_{mn} is the weight associated to sample n and b is a bias parameter. The kernel function considered is a radial basis function. For training the model, i.e. finding optimal values for w_{mn} and b_m, a separate hyperparameter that represents the precision of the corresponding weight is introduced. During the training a significant part of the hyperparameters goes to infinity and the associated weights to zero. Therefore the corresponding sample x_n will play no role in the prediction made by the trained model y_m. The samples for which at the end of the training the weights are non zero are called relevance vectors since they are the samples that are relevant for the prediction made by the obtained model. The training algorithm used is described in detail in (23).

3. EXPERIMENTAL PROCESS

3.1. Data Collection Platform

The robot used in the experiments was equipped with an electronic nose, an ultrasonic anemometer and a laser scanner. The e-nose is an array of five semiconductor gas sensors inserted in an aluminum tube with a fan at one extremity. The purpose of the fan is to damp variations in the airflow to the sensors and therefore limit the influence of the movement of the robot, a summary of the gas sensors used is given in Table 1. Since the substances

chosen for the experiments are heavier than air, whose molecular weight is in average 29 g/mol, the nose was mounted in horizontal position at 10 cm height on the ground, below the laser scanner. The anemometer used to measure the airflow is a Young 81000 Ultrasonic Anemometer and has a resolution of 1 cm/s and a range from 2 cm/s to 40 m/s. Its placement has been a compromise between the need for a measurement of the airflow as close as possible to the nose and to have the measurement as unaffected as possible by the body of the robot. The robot software is based on the Player server (24) that provides easy access to the sensors and the actuators. Moreover, Player provides many high level algorithms such as adaptive Montecarlo localization, wavefront path-planner and vector field histogram obstacle avoidance that were used respectively for localization, global and local path planning.

The experiments have been carried out in a large room (approx. 40 m^2) with no openings. No artificial airflow has been introduced, however according to (12) the flow of the gas is mainly turbulent. Furthermore, in an enclosed space there are still thermodynamic exchanges with the outside and these generate convective airflow. The anemometer is used to capture any dominant airflow that can create an odour plume. Still such a plume is assumed to be characterized by patches and meanders due to the turbulent diffusion. Evidence of these patches can be seen by the fluctuations in the signal response in Figure 1. However without a ground truth it is difficult to create a model of the interaction between the plume and the sensors. Therefore in our approach, the signal segmentation phase attempts to parse the signal into responses due to different patches of gas, where a response to a patch is a rise and a decay as defined in Section 2.2.

The robot was moving following a predefined sweeping trajectory covering an area of roughly 15 m^2. The relative humidity value was varying between 20% and 29.5% and the temperature was varying between 19 °C and 19.5 °C. The experiment run has been repeated executing the sweep with two different orientations (horizontal and vertical) in order to minimize the effects of different path-planning strategies on the classification performances. The gas source used was a 7 x 10 x 4 cm cup placed on the floor adjacent to and outside of the area covered by the movement of the robot. To ensure that the gas does not saturate the environment the cup is covered and the lid is removed 1 minute before the robot starts moving. Between experimental runs the gas source is covered. Fig. 2 provides a graphical representation of the configuration of the experiment together with an example of one of the paths followed by the robot. The average magnitude and direction of the

measured airflow is indicated by the arrow.

The three substances considered in these experiments are ethanol, acetone and isopropyl. In Table 2 we report their chemical formulas and their molecular weight.

3.2. Data Sets

The experiment run has been repeated 15 times with a N-S sweeping trajectory (denoted as a horizontal sweep) and 15 times with a E-W sweeping trajectory (denoted as a vertical sweep), for a total of 30 trials. In total 176 transient responses have been collected. The average airflow measured with the ultrasonic anemometer was 6 cm/s with direction N-E. An example run with horizontal orientation is shown in Fig. 2. From the average wind direction and the position of the responses displayed in the image it is possible to infer the position of the gas plume in the upper part of the middle section of the room.

4. RESULTS

In order to assess which pattern recognition system is suitable for odour classification using continuous sampling a number of factors are evaluated. Firstly, an evaluation of the different techniques used for baseline manipulation, feature extraction and data normalization is made. Secondly, an evaluation of the effects of using a rejection threshold is given. Finally we investigate the performance of the classification algorithm with respect to the distance from the emitting source.

To determine which degree of polynomial should be used to fit our data, as described in Section 2.3, Fig. 3 gives the RMSE for polynomials of degree one to four. In addition the RMSE for the different segments of the signals: rise, decay and the combination of the two, are given. To choose the most appropriate polynomial to fit the signal a tradeoff between the goodness of the fit and the amount of data needed to fit a complex model is taken. Indeed at least $n+1$ measurements are needed to fit an n-th degree polynomial and therefore, by choosing a high degree polynomial, we implicitly set a boundary on the minimum length of the response that the system can process. Note that in the figure there is significant improvement of the RMSE in the third and fourth degree polynomial over the first and second polynomials. This applies for the three different parts of the signal. Also note that the RMSE does not decrease significantly when the degree of the polynomial is

increased from three to four, therefore the third degree polynomial is chosen as an appropriate curve for feature extraction. Consequently the curve fitting method (CF) used in the subsequent tables refers to the third degree polynomial.

The classification performance of the SVM for different combination of baseline manipulation, feature extraction and data normalization are given in Table 3. Only the best 20 combinations are given in the table. In order to evaluate the generalization performance of the classifier, the classification rates in the table have been obtained using a leave-one-out cross-validation (LOOCV). The parameters of the SVM are given in the fifth and sixth column of the table. These parameters have been estimated using a grid search with a 5-fold cross validation. Similar classification results are given in Table 4 for the RVM. A 5-fold cross validation has been used to estimate the width of the kernel that is indicated in the fifth column. Note that the general classification performance for the RVM is slightly lower, however as it will be shown in Fig. 4 the important tradeoff for using a RVM is the added benefit in the tailoring of the classifier through the introduction of a rejection threshold. The advantage is the ability to compensate for the possibility that certain samples may not contain sufficient information to be classified.

An ANOVA analysis was performed in order to evaluate the influence of the different factors, namely signal phase, baseline manipulation, feature extraction and data normalization, on the classification performance. Tables 5 and 6 show the ANOVA tables for the SVM and RVM respectively. The ANOVA tables have been calculated on the results of a 5-fold cross validation. For both the SVM and RVM the Data Normalization, Baseline Manipulation and the combination of the two are the most influential factors (largest F-Values). Indeed the effect of the single factors are much higher than the effect of any combination of them (the combinations that include Data Normalization might have a large F-Value due to the influence of the Data Normalization whose F-Value is two orders of magnitude larger). Considering each of the factors individually and performing a Fisher's-LSD test we obtain the results displayed in Table 7. The table shows the confidence intervals at a confidence level of 0.05 for the mean of the different groups. When the intervals of two techniques are overlapping then no considerable performance difference is observed by using one technique over the other. From the table, for both the SVM and the RVM, the differential baseline manipulation and vector normalization techniques show to be significantly better than the other baseline manipulation and data normalization meth-

ods. The curve fitting method gives the best performance with the SVM but it is inferior to FFT with the RVM. The signal phase is not an influent factor for the SVM nor the RVM classifier.

Fig. 4 displays the classification error and the amount of rejections for a varying threshold obtained using leave-one-out cross validation (LOOCV). For both figures the results obtained using different phases of the signal (rise, decay and rise+decay) are given. The combination of preprocessing methods used are those which resulted in the highest classification rate shown in Table 4 and in Table 7. As expected the amount of rejections increases with the rejection threshold. Interestingly, the classification accuracy improves when the rejection threshold is raised. This suggests that most of the samples that are misclassified have a low posterior probability value. As a result the posterior probability estimation provided by the RVM is an acceptable confidence measure for the classification result.

Finally the last measure of evaluation examines how the classification performance varies with respect to the distance from the emitting source. To compare the two classification methods Fig.s 5 and 6 show a scatterplot of the correctly classified responses for the SVM and for varying rejection threshold for the RVM. The scatterplots give a spatial representation of the classification performance where the location of the odour source is marked by the square. Correctly classified responses are indicated by the smaller circles whereas incorrectly classified responses are indicated by red triangles. Note in the figure that successfull classification is not only obtained in close proximity to the source but also at distances of approximately 3 meters from the source. Also in Fig. 6 the rejection threshold of the RVM can be used to minimize classification errors, as almost no incorrectly classified responses are found with the threshold above 0.95. Indeed lowering the threshold to 0.33 implies no rejections (as there are only three classes) and the number of incorrectly classified responses is much higher. In fact these results are comparable to the SVM in Fig. 5. Table 8 gives the average distance of the correctly classified responses and the incorrectly classified responses in Fig. 6 with a rejection threshold of 0.33. Another interesting feature to note from these figures is that the misclassified responses are not necessarily concentrated far from the location of the gas source. This suggests that it is not only possible to use a continuous sampling method but also to use this method in an open environment where the position of the gas source is not known in advance. To prove this more experimentation is necessary, however the 2-sample t-test performed on the presented results rejected the

hypothesis that the two samples come from distribution with different mean at 0.95 significance level.

5. CONCLUSIONS AND FUTURE WORKS

This paper has investigated the possibility to use a new sampling method for odour classification where the sensor array is exposed and continuously samples the environment. The analysis has been carried out starting from the works that have been done for three-phase sampling based devices and trying to understand to which extent their methods can be applied. Due to the differences in the properties of the signal when compared to a signal sampled using the three-phase technique, a segmentation step is introduced, which isolates the relevant parts of the signal that can then be passed to the pattern recognition algorithm. Since the segmentation policy proposed is not assured to be optimal, estimation of posterior probabilities and a rejection class has been introduced. This opens the possibility to discard samples for which the classifier is undecided.

The results presented constitute an important step towards reliable classification for artificial olfaction applications such as continuous online pollution monitoring. As future work we would like to consider the extension to the problem of classification in environments where more than one substance is present. Another interesting development would be the introduction of committees or techniques for combining classifiers. This would allow the exploitation of the differences in the behavior of the different pattern recognition algorithms and possibly increase the classification performance. Also the introduction of a feature selection stage would carry a two folded benefit: it would shrink the size of the model making it more suitable for the deployment on a small embedded system, and, if the features are localized in time like the DWT, it would provide more insight on which part of the signal is most relevant for classification.

References

[1] Schweizer-Berberich M., Vaihinger S., and Göpel W. Characterisation of food freshness with sensor arrays. *Sensors and Actuators B*, 18-19:282–290, 1994.

[2] Achim J. Lilienthal, Amy Loutfi, and Tom Duckett. Airborne chemical sensing with mobile robots. *Sensors*, 6:1616–1678, October 2006.

[3] H. Ishida, T. Nakamoto, and T. Moriizumi. Remote sensing of gas/odor source location and concentration distribution using mobile system. *Sensors and Actuators B*, 49:52–57, 1998.

[4] P. Pyk, S. Bermúdez Badia, U. Bernardet, P. Knüsel, M. Carlsson, J. Gu, E. Chanie, B. S. Hansson, T. C. Pearce, and P. F. Verschure. An Artificial Moth: Chemical Source Localization Using a Robot Based Neuronal Model of Moth Optomotor Anemotactic Search. *Auton Robot*, 20:197–213, 2006.

[5] A. Lilienthal, H. Ulmer, H. Fröhlich, A. Stützle, F. Werner, and A. Zell. Gas source declaration with a mobile robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2004)*, pages 1430 – 1435, 2004.

[6] Thomas Eklov, Per Martensson, and Ingemar Lundstrom. Enhanced selectivity of mosfet gas sensors by systematical analysis of transient parameters. *Analytica Chimica Acta*, 353:291–300, 1997.

[7] R. Gutierrez-Osuna, H. T. Nagle, and S. S. Schiffman. Transient response analysis of an electronic nose using multi-exponential models. *Sensors and Actuators B: Chemical*, 61:170–182, 1999.

[8] L. Carmel, S. Levy, D. Lancet, and D. Harel. Transient response analysis for temperature modulated chemoresistors. *Sensors and Actuators B: Chemical*, 93(1-3):67–76, 2003.

[9] C. Distante, M. Leo, P. Siciliano, and K.C Persaud. On the study of feature extraction methods for an electronic nose. *Sensors and Actuators B: Chemical*, 87:274–288, 2002.

[10] D. Martinez, O. Rochel, and E. Hughes. A biomimetic robot for tracking specific odors in turbulent plumes. *Autonomous Robots*, 20:185–195(11), June 2006.

[11] Amy Loutfi, Silvia Coradeschi, Achim J. Lilienthal, and Javier Gonzalez. Gas distribution mapping of multiple odour sources using a mobile robot. *Robotica, To Appear*, 20, 2008.

[12] P. J. W. Roberts and D. R. Webster. Turbulent Diffusion. In H. Shen, A. Cheng, K.-H. Wang, M.H. Teng, and C. Liu, editors, *Environmental Fluid Mechanics - Theories and Application*. ASCE Press, Reston, Virginia, 2002.

[13] Julian W. Gardner and Philip N. Bartlett. *Electronic Noses: Principles and Applications*. Oxford Science Publications, 1991.

[14] R. Gutierrez-Osuna and H. T. Nagle. A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 29:626–632, 1999.

[15] R. Gutierrez-Osuna, A. Gutierrez-Galvez, and N. U. Powar. Transient response analysis for temperature modulated chemoresistors. *Sensors and Actuators B: Chemical*, 93(1-3):57–66, 2003.

[16] P. Duhamel and M. Vetterli. Fast fourier transforms: A tutorial review and a state of the art. *Signal Processing*, 19:259–299, 1990.

[17] Mallat S. *A wavelet tour of signal processing*. Academic Press, 1998.

[18] Amine Bermak, Sofian Brahim Belhouari, Minghua Shi, and Dominique Martinez. Pattern recognition techniques for odor discrimination in gas sensor array. *Encyclopedia of Sensors*, X:1–17, October 2006.

[19] Chih C. Chang and Chih J. Lin. *LIBSVM: a library for support vector machines*, 2001.

[20] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[21] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.

[22] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[23] M. E. Tipping. and A. C. Faul. Fast marginal likelihood maximisation for sparse Bayesian models. *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003.

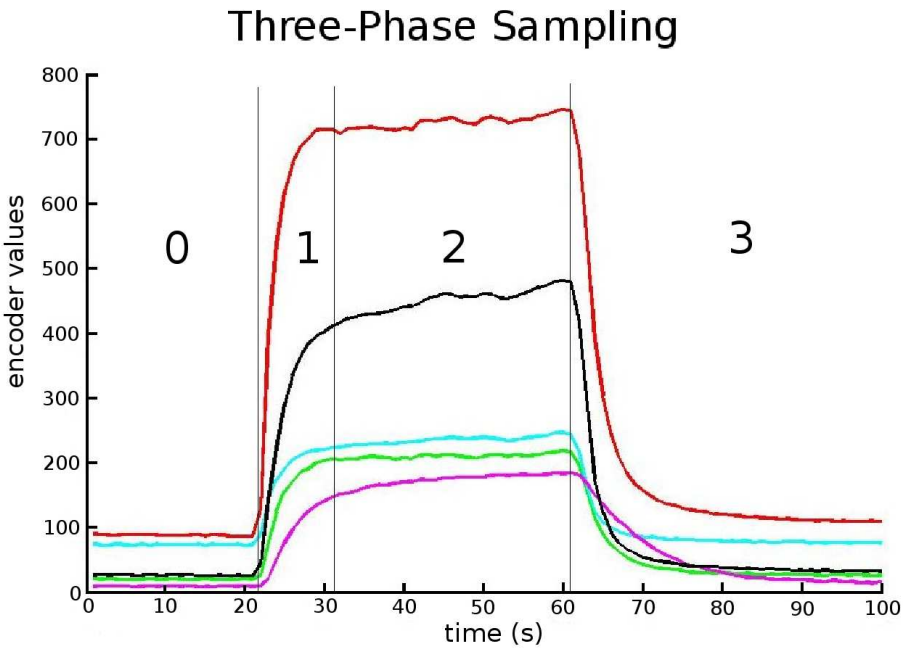
[24] B. Gerkey, R. T. Vaughan, and A. Howard. The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems. In *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR)*, pages 317–323, 2003.

Marco Trincavelli was born in Italy on March 24, 1981. He received both the BSc degree (2003) and the MSc degree (2006) in Computer Science Engineering from the Politecnico di Milano, Milan, Italy. He additionally received a MSc degree in Electrical Engineering and Computer Science from the Lund Tekniska Högskola, Lund, Sweden, in 2006. In 2005-2006 he worked as a consultant for Digitec srl. Lecco, Italy working mostly in the area of computer vision. Since spring 2006 he has been a graduate student at the Center for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden. His current research includes the study of artificial olfaction for mobile robots with particular focus on the discrimination of odours.

Silvia Coradeschi is an associate professor at Örebro University, Sweden. She has received a MSc in Philosophy from the University of Florence, a MSc in Computer Science from the University of Pisa and a Ph.D. in Computer Science at Linköping University. Her main research interest is in establishing the connection (anchoring) between the symbols used to perform abstract reasoning and the physical entities, which these symbols refer to. She also works in co-operative robotics, mobile robotics and artificial olfaction applied to robotics.

Amy Loutfi is a post-doctoral researcher at Örebro University, Sweden. She has received a BSc. in Electrical Engineering from the University Of New Brunswick, Canada and a PhD. in Computer Science from Örebro University. She has several years experience with electronic nose devices and her main research interest is in the integration of electronic nose into intelligent systems such as mobile robots. She also is interested in mobile robotics for environmental monitoring, artificial intelligence, and distributed robotic systems.

List of Tables



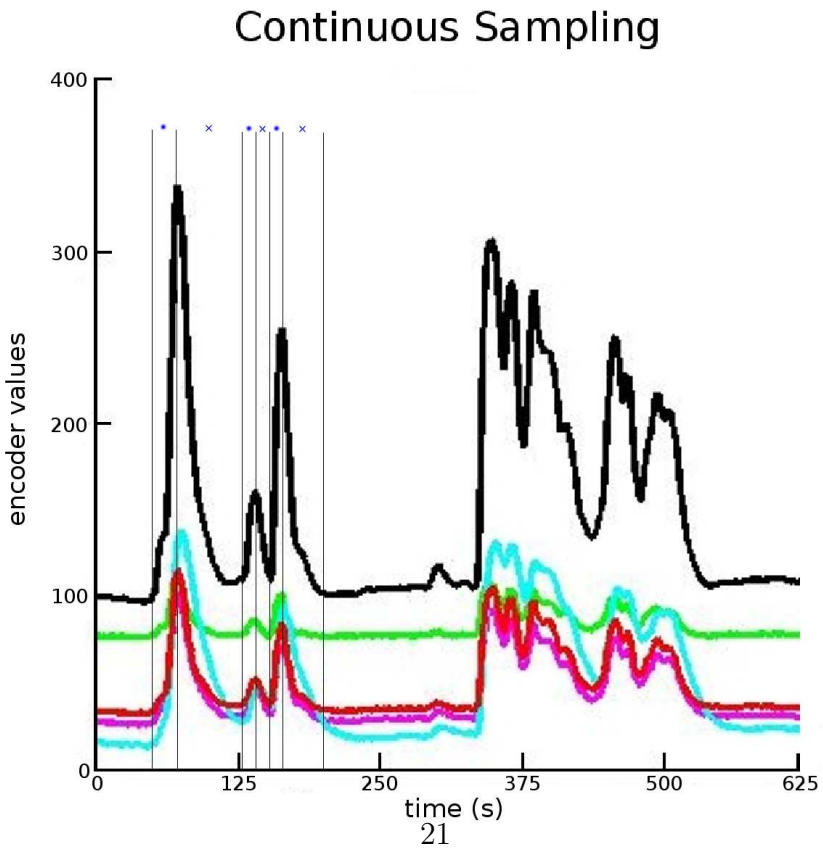


Figure 1: Upper: example of signal collected using the traditional three phases sampling process. The phases of the sampling have been numbered in the following way: 0 baseline, 1 transient, 2 steady state, 3 recovery. Lower: example of signal collected with the mobile robot in an uncontrolled environment. The first part of the response shows the segmentation. The asterisks indicate the parts of the signal that have been segmented as rise phase, while the diagonal crosses indicate the recovery phases. Both figures indicate the response from the five gas sensors described in Table 1

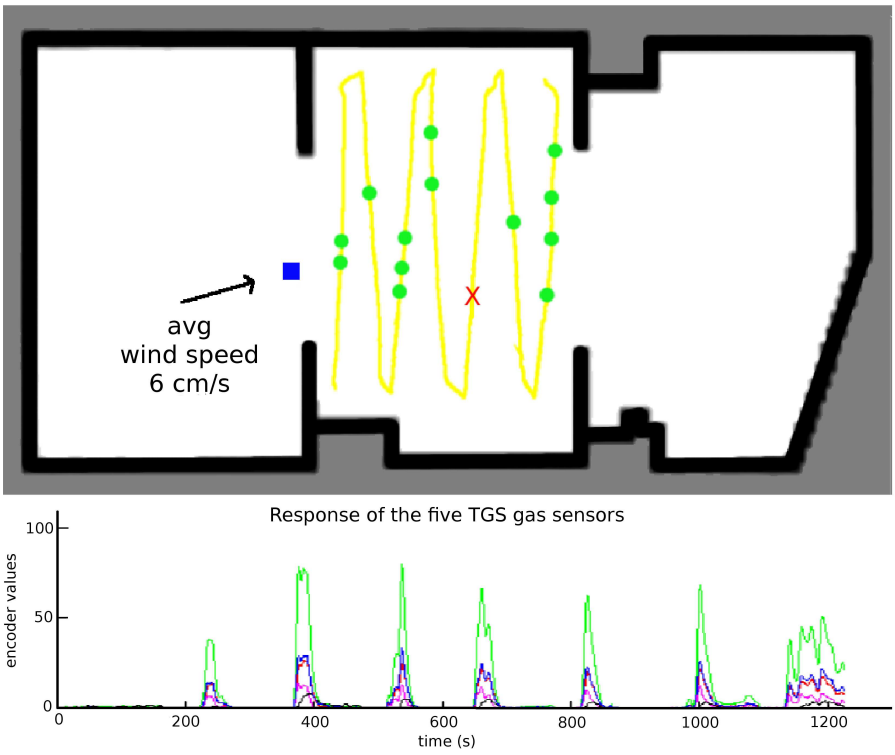


Figure 2: Upper: Example of experiment using isopropyl. The robot follows a horizontal sweeping trajectory frequently entering and exiting the plume. The arrow shows the average direction and magnitude of the windflow. The square indicates the position of the source. The solid line indicates the trajectory of the robot. The circles and crosses are locations in which was obtained a transient that could be classified having the rejection threshold set at 0.8. Lower: The actual readings collected during the run by the sensors in the electronic nose.

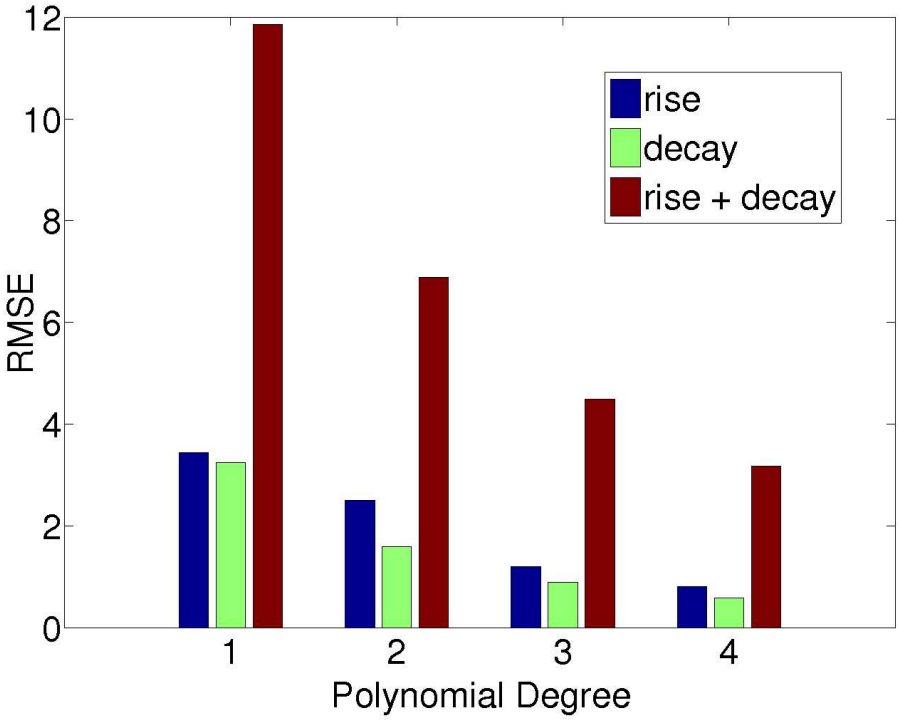


Figure 3: RMSE for polynomial of different degree. The curves are fitted to the different parts of the signal, namely rise, decay and both together.

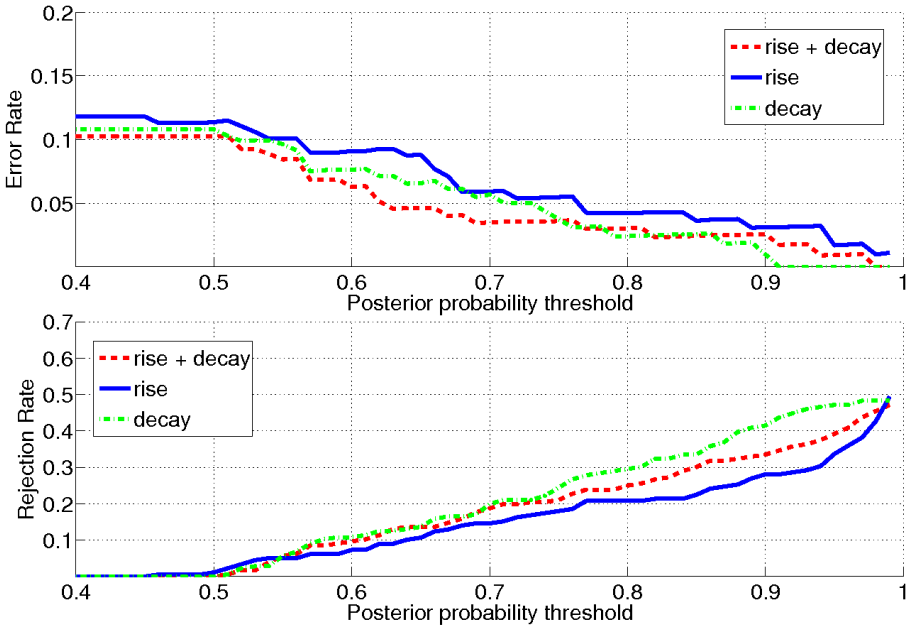


Figure 4: Error and rejection rate for the RVM classifier with a varying rejection threshold. The three lines represent the performance of classifier trained with different parts of the signal: rise, decay and both together.

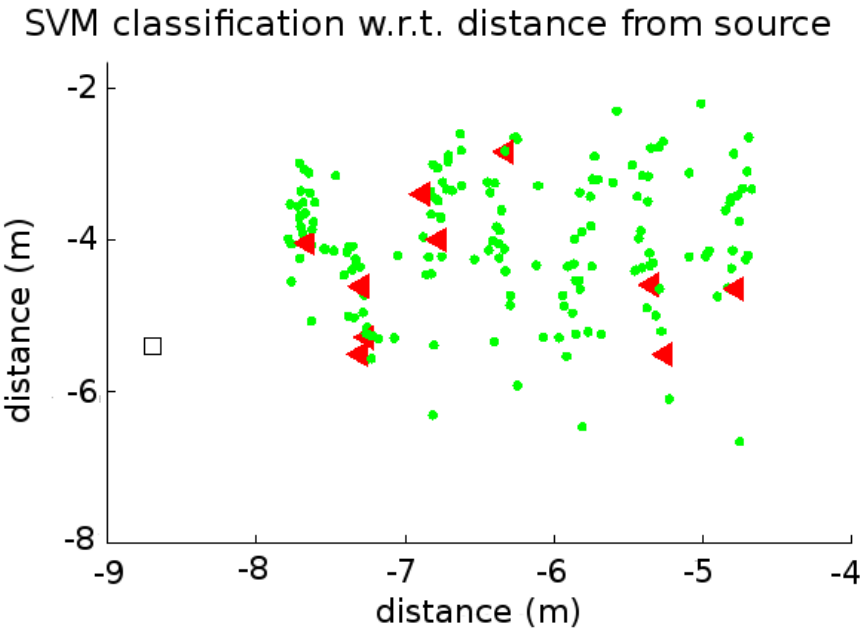


Figure 5: Scatterplot of the sensor responses classified with the SVM. The circles represent correctly classified responses, while the triangles are the errors. The distances on the axes are expressed in meters.

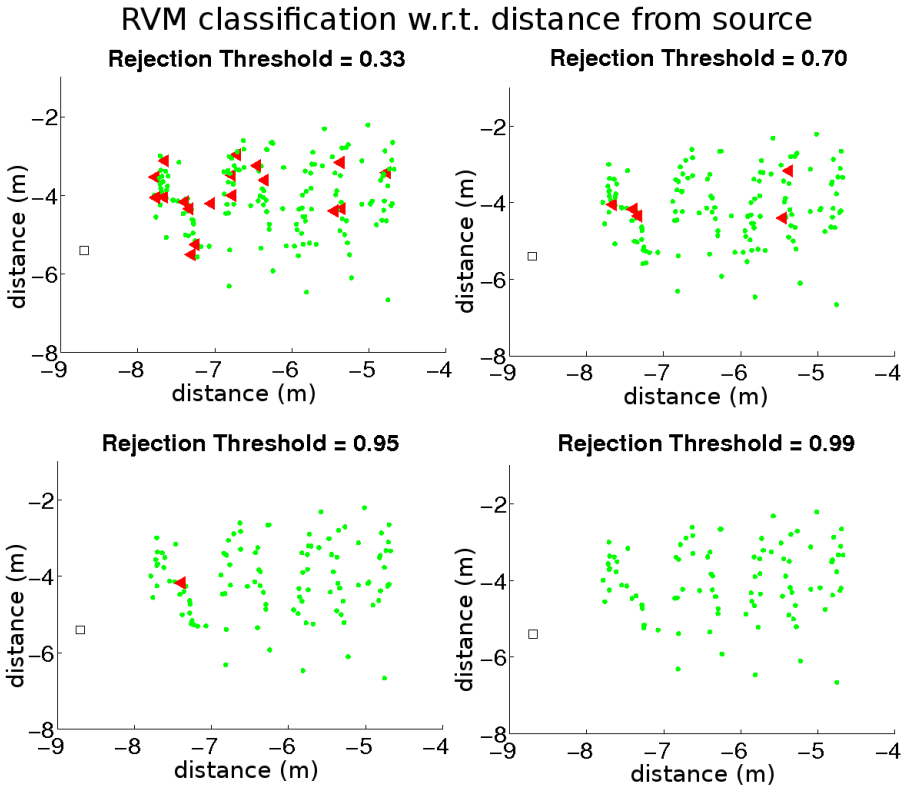


Figure 6: Scatterplot of the sensor responses classified with the RVM. The circles represent correctly classified responses, while the triangles are the errors. The distances on the axes are expressed in meters.

Model	Gases Detected	Quantity
Figaro TGS 2600	Hydrogen, Carbon Monoxide	2
Figaro TGS 2602	Ammonia, Hydrogen Sulfide, VOC (volatile organic compound)	1
Figaro TGS 2611	Methane	1
Figaro TGS 2620	Organic Solvents	1

Table 1: Gas sensors in the electronic nose and their respective target gases.

Substance	**Chemical Formula**	**Molar Mass**
Ethanol	C_2H_5OH	46 g/mol
Acetone	CH_3COCH_3	58 g/mol
Isopropyl	C_3H_8O	60 g/mol

Table 2: List of substances used in the experiment

Segmented Phase	**Baseline Manipulation**	**Feature Extraction**	**Data Normalization**	**Parameter γ**	**Parameter C**	**Classification Rate**
R & D	DIFF	CF	VN	2^{-4}	2^{15}	94.3%
R	DIFF	FFT	VAS	2^{-3}	2^{11}	94.3%
R & D	REL	CF	VN	2^{-9}	2^{23}	93.8%
R & D	DIFF	CF	DAS	2^{-7}	2^{25}	92.7%
D	DIFF	CF	DAS	2^{-6}	2^{18}	92.0%
D	DIFF	FFT	VN	2	2^{7}	91.5%
D	DIFF	FFT	DAS	2^{-8}	2^{7}	90.9%
D	REL	FFT	VN	2^{-3}	2^{16}	89.9%
D	DIFF	CF	VN	2^{-3}	2^{25}	89.8%
D	DIFF	DWT	VN	2	2^{9}	89.8%
D	FRAC	FFT	VN	2^{-4}	2^{13}	89.7%
R	FRAC	FFT	VN	2^{2}	2^{6}	89.7%
R & D	DIFF	DWT	DAS	2^{-9}	2^{25}	89.2%
R	DIFF	CF	VN	2^{-10}	2^{24}	89.2%
R & D	DIFF	DWT	VAS	2^{-8}	2^{25}	88.6%
D	DIFF	DWT	VAS	2^{2}	2^{7}	88.6%
R	DIFF	DWT	VAS	2^{-10}	2^{18}	88.6%
R	DIFF	DWT	VN	2^{-10}	2^{23}	88.6%
R	DIFF	FFT	DAS	2^{-10}	2^{25}	88.6%
R & D	FRAC	CF	VAS	2^{-7}	2^{25}	88.6%

Table 3: Classification Performance and parameters used for the pattern recognition algorithm based on the Support Vector Machine Classifier. Classification rate has been evaluated using LOOCV

Segmented Phase	**Baseline Manipulation**	**Feature Extraction**	**Data Normalization**	**Parameter γ**	**Classification Rate**
D	DIFF	FFT	VN	2^{-1}	84.7%
D	FRAC	DWT	VAS	2^{-2}	84.6%
D	FRAC	DWT	VN	2^{-1}	84.6%
D	FRAC	CF	VAS	2^{3}	84.0%
R & D	FRAC	DWT	VN	2^{4}	83.4%
R	REL	FFT	VAS	2^{4}	83.1%
D	DIFF	DWT	VN	2^{5}	83.0%
R & D	FRAC	FFT	VN	2^{4}	82.9%
R & D	REL	FFT	VAS	2^{4}	82.6%
D	FRAC	FFT	VN	2	82.3%
R & D	REL	FFT	VN	2^{-1}	82.0%
D	FRAC	CF	VN	2	81.7%
D	FRAC	FFT	VAS	2	81.7%
R	FRAC	FFT	VN	2	81.7%
R & D	REL	DWT	VN	2	81.5%
D	REL	DWT	VN	1	81.5%
D	DIFF	DWT	VAS	1	81.2%
R	DIFF	FFT	VN	1	81.2%
D	REL	CF	VN	2^{-1}	80.9%
D	REL	FFT	VN	2^{-1}	80.9%

Table 4: Classification Performance and parameters used for the pattern recognition algorithm based on the Relevance Vector Machine Classifier. Classification rate has been evaluated using LOOCV

Factor	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F - Value**
Signal Phase (SP)	112.4	2	56.2	1.24
Baseline Manipulation (BM)	6702.9	2	3351.5	73.73
Feature Extraction (FE)	1006.1	2	503.1	11.07
Data Normalization (DN)	21418.8	2	10709.4	235.59
SP x BM	41.3	4	10.3	0.23
SP x FE	624.2	4	156	3.43
SP x DN	327.7	4	81.9	1.8
BM x FE	1070	4	267.5	5.88
BM x DN	8617	4	2154.2	47.39
FE x DN	3537.5	4	884.4	19.45
SP x BM x FE	579.8	8	72.5	1.59
SP x BM x DN	102.1	8	12.8	0.28
SP x FE x DN	728.3	8	91	2
BM x FE x DN	2029.7	8	253.7	5.58
SP x BM x FE x DN	751.4	16	47	1.03
Error	14728.3	324	45.5	
Total	62377.5	404		

Table 5: ANOVA table for the SVM classifier

Factor	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F - Value**
Signal Phase (SP)	591.5	2	295.7	4.59
Baseline Manipulation (BM)	625.7	2	312.9	4.85
Feature Extraction (FE)	773.1	2	386.6	6.0
Data Normalization (DN)	35154.1	2	17577.1	272.67
SP x BM	784.3	4	196.1	3.04
SP x FE	523.5	4	130.9	2.03
SP x DN	510.3	4	127.6	1.98
BM x FE	147.1	4	36.8	0.57
BM x DN	3989.8	4	997.4	15.47
FE x DN	1444.8	4	361.2	5.6
SP x BM x FE	586.5	8	73.3	1.14
SP x BM x DN	576.5	8	72.1	1.12
SP x FE x DN	753.4	8	94.2	1.46
BM x FE x DN	865.8	8	108.2	1.68
SP x BM x FE x DN	1106	16	69.1	1.07
Error	20886.2	324	64.5	
Total	69318.7	404		

Table 6: ANOVA table for the RVM classifier

Signal Phase		
	SVM	RVM
Rise	78.09 - 79.71	70.17 - 72.09
Decay	79.37 - 80.99	71.79 - 73.71
Rise & Decay	78.87 - 80.49	68.83 - 70.75
Baseline Manipulation		
	SVM	RVM
Differential	84.52 - 86.13	71.91 - 73.83
Fractional	75.53 - 77.14	69.96 - 71.88
Relative	76.29 - 77.90	68.91 - 70.84
Feature Extraction		
	SVM	RVM
DWT	77.90 - 79.52	70.61 - 72.53
FFT	77.44 - 79.06	71.76 - 73.68
Curve Fitting	80.99 - 82.61	68.42 - 70.35
Data Normalization		
	SVM	RVM
Vector Normalization	85.66 - 87.28	79.41 - 81.33
Vector Autoscaling	81.96 - 83.57	73.9 - 75.83
Dimension Autoscaling	68.72 - 70.33	56.47 - 59.40

Table 7: Results of the LSD Fisher's test.

	Average	**Standard Deviation**
Correct Classification	2.93 m	0.93 m
Erroneous Classification	2.58 m	0.87 m

Table 8: Average and standard deviation of the distance from the gas source for classified transients. The rejection threshold is set to zero, thus all transients have been classified.