Asimina KARAGEORGIOU

Master Micro and Nano Technologies for Integrated Systems – NANOTECH 2017

IBM ALMADEN - 650 Harry Rd, San Jose, CA 95120, USA

Characterization and data analysis of commercial gas sensors

From 26/02/2017 to 25/08/2017

Under the supervision of:

- **Company supervisor: Andrea, FASOLI, andrea.fasoli@ibm.com**
- **Phelma Tutor: Youla, MORFOULI, email address**

Confidentiality: no

Ecole nationale
supérieure de physique,
électronique, matériaux

Phelma
Bât. Grenoble INP - Minatec
3 Parvis Louis Néel - CS 50257
F-38016 Grenoble Cedex 01

Tél +33 (0)4 56 52 91 00
Fax +33 (0)4 56 52 91 03

http://phelma.grenoble-inp.fr

Acknowledgements

I feel the need to express my gratitude to my mentor, Andrea Fasoli, for trusting me to undertake certain responsibilities in the Sensor Lab of IBM, for organizing the project and for the guidance throughout the completion of the thesis.

I thank Aminat Adebiyi, who helped me make my first steps in the science of programming and being friendly during our common working sessions.

I thank Luisa Bozano, for the effort she put in order to make all interns of IBM feel satisfied project-wise and life-wise during their respective internships.

Last but not least, I thank the coordinators of the master Nanotech for giving me the opportunity to conduct my master thesis in the Research Center of IBM Almaden, in the first place.

Summary

The present report is the description of the 6-month internship at the IBM Almaden research center. The research that has been conducted is related to the sensor project, which started after the company has been assigned to optimize the performance of the gas sensor with respect to those which already exist on the market. A scientific goal like this one demands primarily the investigation and the understanding of the commercial sensors that already exist on the market. However, the deeper understanding of the commercial sensors inspired the sensor group to design a board made of commercial sensors that could operate as an electronic nose. Subsequently, the second goal of the company involves the selection of the most suitable commercial sensors after evaluating their performance for the design of that board. For this reason the current report is related to 1. the description of the experimental set-up that has been used in order to examine a series of commercial sensors 2. their respective experimental results 3. the description of an automatic process that has been established with a programming language for what concerns a faster data analysis of the commercial sensors and 4. the introduction of machine learning approaches, since a cognitive part usually accompanies the electronic noses for the determination of unknown inputs in terms of the nature of the gas and its respective concentration.

Ce rapport est la description du stage de six mois au centre de recherche IBM Almaden. Les études qui ont été conduites sont liés au « sensor project », qui a commencé après que l'entreprise ait été assigné à l'optimisation des performances de capteurs de gaz par rapport à ceux présent sur le marché. Un effort de recherche comme celui-ci repose essentiellement sur une investigation et une compréhension des capteurs commerciaux existant sur le marché. Cependant, une compréhension profonde des capteurs commerciaux ont inspiré le « sensor group » à confectionner une plaque composée de capteurs commerciaux qui agit comme un nez électronique. Le second but impliques la sélection des capteurs les plus approprié après avoir évalué leurs performances. Pour cette raison, ce rapport correspond en 1. La description du dispositif expérimental qui a été utilisé pour examiner une série de capteurs commerciaux. En 2. Les résultats expérimentaux de ces derniers seront présentés. Le 3. est la description du processus automatique qui a été établi utilisant un langage de programmation pour ce qui concerne une analyse plus rapide des capteurs commerciaux. Enfin en 4. sera introduit une approche de Machine Learning, car une partie cognitive accompagne le nez électronique pour la détermination des entrées inconnues en termes de nature de gaz et de concentration.

Il presente report è la descrizione dell'internship semestrale svoltosi presso il centro di Ricerca Almaden - IBM. La Ricerca condotta è inerente il progetto di un sensore, con lo scopo di ottimizzare le prestazioni di sensori di gas, rispetto ai già esistenti nel mercato. Il primo obiettivo scientifico richiesto riguarda l'investigazione e lo studio del funzionamento dei sensori di gas già esistenti nel mercato. Tuttavia, uno studio più approfondito dei sensori commerciali ha permesso al gruppo di sviluppare un innovativo naso elettronico. Il secondo obiettivo riguarda la scelta dei più inerenti sensori, dopo aver valutato le loro prestazioni. Per queste motivazioni, il presente report è inerente: 1. la descrizione del set-up sperimentale che è stato utilizzato, al fine di esaminare una serie di sensori commerciali 2. le misure sperimentali relative 3. la descrizione di un processo automatico che ha permesso di stabilire, tramite linguaggio di programmazione, una più veloce analisi dei dati, nell'ambito dei sensori commerciali 4. l'introduzione dell'approccio "machine learning" , dal momento che una parte del processo cognitivo di solito accompagna i nasi elettronici per la determinazione di input sconosciuti, in termini della natura del gas presente e delle sua rispettiva concentrazione.

Contents

1.1 INTRODUCTION

The work of the present thesis has been conducted in IBM Almaden Research Center, where I worked as a student research assistant, for a period of six months (~ February to August 2017).

Metal Oxide-based gas sensing is a mature technology that has been investigated for the past several decades, the first commercial devices dating back to the 1960s. To date, this technology enables a wide range of sensing applications in diverse fields such as industrial chemical safety, environmental monitoring, quality control, airport security, and healthcare. However, the fundamental principles of operation of Metal Oxide materials are not yet totally understood by the scientific community and further improvements in sensor performances appears to be possible. This realization, in combination with the tremendous opportunities presented by the blooming of the Internet of Things paradigm, provide the drive to develop material-wise optimized sensors that detect gases on a real-time basis with the support of machine learning techniques.

Over the course of the past few years, IBM Almaden has focused its attention on the field of gas sensing in order to bring to the market a new generation of sensors and gas sensing solutions. In the frame of this scientific aim a series of parallel investigations has been going on concerning topics such as: 1) characterization, reverse engineering and understanding of existing commercial sensors 2) characterization of home-made sensors in order to achieve a sensitivity-stability trade-off by optimizing the materials and the fabrication conditions 3) construction of a board whose purpose is the gas and concentration discrimination due to careful selection of commercial sensors mounted on the very same board and 4) machine learning and data analysis to support the cognitive part of the above mentioned projects.

The present report relates to the last three of these topics, i.e characterization of commercial sensors followed by the data analysis of the results in order to select the suitable sensors for the construction of the respective board. Early steps on the machine learning side have also been realized.

In Figure 1 one can appreciate the implementation of the project on a timescale. It can be deducted that for the completion of the project a number of technical skills was mandatory to be acquired and were related to the following: performing electrical measurements, python programming language and octave programming language. Specifically, during the first period of the internship, effort and time was consumed mainly in performing measurements, obtaining data and also software skills that were necessary for the subsequent analysis of the respective data.

The electrical characterization was performed for eighteen commercial gas sensors, the average price of which is $12.70 (total cost: $228.60), using the following instrumentation: 1. Owlstone OFC flow controller, OHG humidity generator, OVG-4 callibration gas generator (total cost: $1600) and 2. Vice Metronics model 230-15a-RD (total cost: $748). Two extra mass flow controllers were also necessary for the completion of the experimental set-up and specifiacally, the GM50A/GM51A mass flow controllers by MKS Instruments (total cost: $400). Given the fact that the boards on which the commercial sensors were mounted in order to perform the electrical measurements (chapter 3.1) and the chamber, were manufactured in house, the respective cost is not to be evaluated. Therefore, the total cost of the project is appreciated to be $2.976.

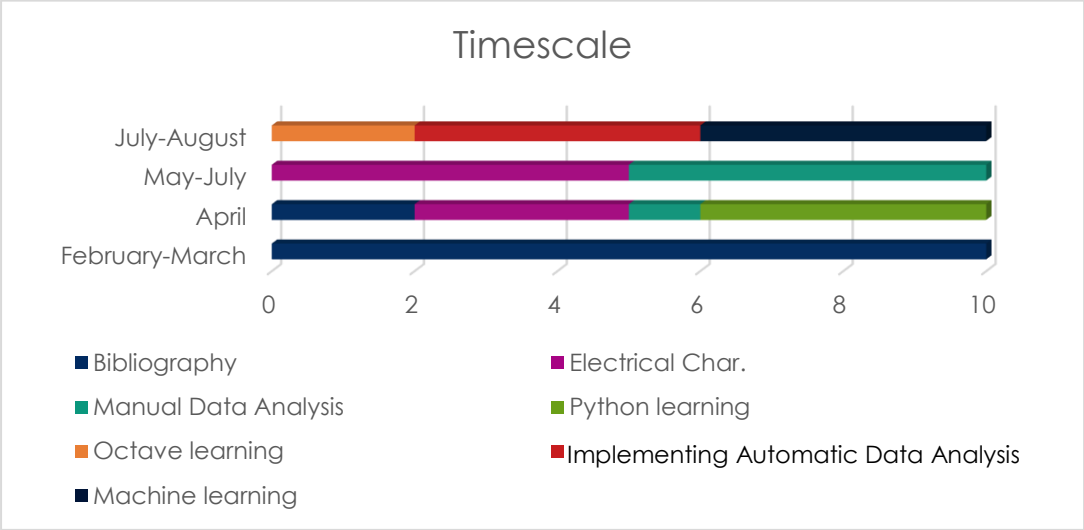


Figure 1 : Timescale of project implementation

1.2 IBM ALMADEN CALIFORNIA

The present work has been conducted in IBM Research – Almaden, which is positioned in Silicon Valley, in San Jose, California and is one of IBM's twelve worldwide research labs that contribute to the research and development department of IBM. The company itself holds a series of admirable rankings on the Forbes list such as: #75 in Sales , #25 in Profit, #245 in Assets, #44 in Market value , #13 World's Most Valuable Brands (1), while the recent annual revenues of the company can be appreciated in Figure 2.

The Almaden Center focuses its attention on 1) Science and Technology, including fundamental science, nanotechnology, spin physics and photoresists 2) Cognitive Solutions & Foundations such as content management, human-computer interaction, text analytics, services-oriented architectures and healthcare informatics 3) Accelerated Discovery for Industry Solutions focusing on large scale, people and information-intensive challenges and 4) Global Storage Systems from storage and file systems to server software and systems management (2).

The content of this thesis falls in the first category of Science and Technology and most specifically it is included in the sensor project of the company, which is increasingly gaining more and more interest. The research group –of which the author of this thesis was ultimately considered to be a member for the given time of her thesis completion - that has been working on the specific project, was realizing weekly meetings in order to present the latest progress, challenges of the work, and to brainstorm for the future direction of the project.

Subsequently, the well–structured environment and well-organized schedules constitute excellent features for any intern that has the opportunity to pursue her thesis in this company, since they constitute a tremendous starting point for career which complies with appropriate procedures.

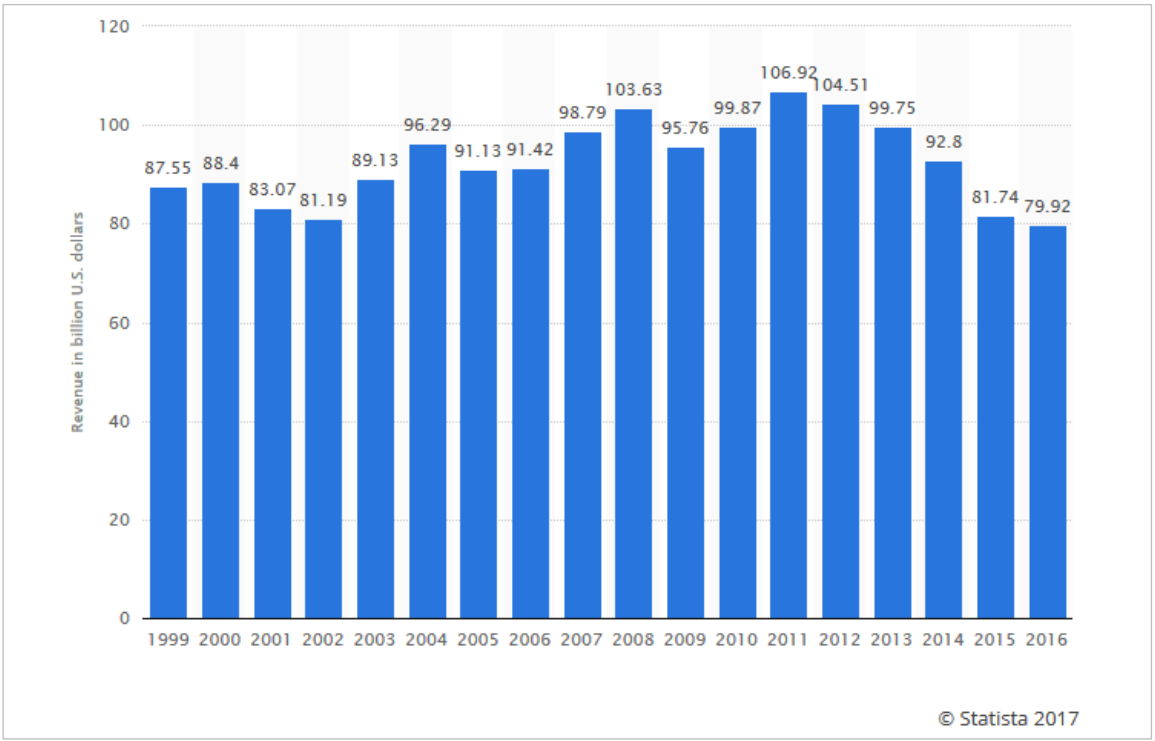


Figure 2: Annual Revenue of IBM in U.S. dollars **(2)**

1.3 BIBLIOGRAPHY

(1). Retrieved from https://www.research.ibm.com/labs/almaden/

(2). Retrieved from https://www.forbes.com/companies/ibm/

2.1 THEORETICAL PART

The electronic nose resulted from the initial idea to manufacture a device that mimics the olfactory system for smells (1). Even though the specific device concept was proposed over

twenty years ago, the ambitious goal has not yet been reached (2). Amongst the main challenges that this field faces are sensitivity, specificity and stability of the individual sensors. In addition, while the electronic nose can be designed for detection of a certain range of gases of interest, those gases are often part of a complex background that acts as interference. Environmental and device conditions such as the partial oxygen pressure at the surface, the local surface temperature may also act as interfering factors. In other words, it can be understood that there is a need for more sensitive, stable sensors and for pattern recognition approaches in order to achieve the determination of the gas input.

An electronic nose comprises an array of different partially-specific sensors. One of the efforts put by IBM's sensor-research-group is related to the assembly of a Metal-oxide-sensors (MOS sensors) array. Specifically, the electrical measurements and the data analysis that have been conducted in the frame of this project aim to the selection of the four more suitable MOS sensors for the construction of a meaningful MOS array. For this reason, in this chapter the following parts are intended to be covered: the principles of operation of the electronic nose, its future trends, understanding the physics behind the operation of individual MOS sensors, and some state of the art data analysis and classification approaches.

2.1.1 Electronic nose: applications, challenges and future trend

Electronic noses rely on arrays of sensors. As far as conventional sensors are concerned, specificity and reversibility constitute an important trade-off one needs to deal with. Specificity refers to the ability of a sensor to discriminate between the analyte of interest and interferences-gases. Reversibility is related to the successful adsorption of the gas molecules from the sensing film when the gas is gone. Therefore, their trade-off is the result of the fact that since a strong specificity demands strong binding between the gas molecules and the film, it becomes more difficult for them to adsorb on a later step. When it comes to the human olfactory system, the corresponding smell receptors have a short lifetime (a few weeks), in order to retain specificity and reversibility (3). However, such short lifetime is prohibitive technologically and financially speaking. In that sense, instead of replacing the sensors – which would be the equivalent of the birth of a new human smell receptor, efforts have been noted in order to tune and improve the specificity of the MOS sensors with alternative ways.

The tuning of the specificity has been attempted by 1) using different sensing materials of the MOS sensor, 2) using dopants in the sensing film itself, 3) realizing different morphologies (amorphous or crystalline) of the film, 4) printing different shapes for electrodes and 5) trying different operating temperatures (See chapter 2.1.2 for better understanding) (4). Approaches 2-5 are more straightforward, since they are related to nanotechnology fabrication and design steps that are well established in the field. The realization of different sensing films however may also follow more exotic approaches. For example, there have been reports claiming that imprinted molecules receptors, DNA, or even whole cells may increase the sensor sensitivity and specificity (5-8).

In addition to fabrication-related solutions, sensitivity and selectivity may also be tuned in ways that correspond to sample pretreatment solutions. The idea lies in the fact that by applying suitable filters, the background gases will be filtered out and as a result that will reduce their interference effect and hence will increase selectivity and sensitivity. This solution doesn't provide only beneficial effects though, since the extra filters increase the complexity, the cost and the physical dimensions of the device (9).

As previously mentioned, several factors may deteriorate the performance of the MOS sensors such as the background gases, the partial oxygen pressure at the surface, the local surface temperature, the fluctuations of humidity and the instability of the sensors (specifically the drift of the baseline which is going to be mentioned in chapter 2.1.2). The effects that occur from all those factors result in fluctuations of the output signal of the sensors and since their intensity vary in an unpredicted way it becomes difficult and sometimes even impossible to include respective corrections when it comes to the calibration of the sensors. The solution to problems like that may lie to the construction of adaptive neural networks that can process the input data in an appropriate way and classify the gases with high accuracy regardless of such interferences. The explanation of the neural network will take place in chapter 2.1.4.

Other strategies consist in combing different types of sensors, each one of which could still undergo the above mentioned fabrication modifications. Surface or bulk acoustic or wave (SAW, BAW) sensors, metal oxide field effect transistors (MOSFETs), or conducting polymer (CP) sensors are suitable candidates to be integrated in the same array.

At this point it should be remarked however that solutions of this type (combining sensors) puts at risk the quality of the output signal, since the noise of the signal and the dimension of the respective array constitute an additional trade-off, when it comes to designing an electronic nose (10). Based on the latter, it becomes obvious that accurate selection of the sensors comprising an array plays a crucial role: the number of sensors should be the minimum number of sensors needed in order to retrieve valuable information at the output, without unnecessarily increasing the physical dimensions, nor the noise, of the device. The latter challenge has been confronted by selecting sensors based on a specific application (the kind of gases the nose is intended to sense) and by relying on knowledge of existing analytical data, in order to choose the best combination of sensors that provides the maximum sensitivity to gases of interest as well as a set of linearly independent outputs.

The different kind of sensors that are mentioned above do constitute a more classical electronic nose. Mobility spectrometers-based sensors and flash chromatographs-based sensors are less conventional ways to detect and determine the sample-gas. As it will be discussed in the following chapter, the more classical sensor-based electronic nose gives an output based on input features, which are related to the electrical signals that the gas produces after reacting with the sensing film of the sensor. The less conventional electronic noses do rely on features, which are related to the time-of-flight, retention time or mass ratios of the gas components. This generation of electronic noses is not necessarily more suitable or of

better performance when compared to the traditional electronic noses. The choice of the most suitable kind of electronic nose seems to be related on the application itself (11-14).

Based on the above mentioned remark, it can be understood that it is indeed meaningful that the present thesis focuses on the research of the more classical sensor-based electronic nose. It seems that there is still room of improvement even when it comes to more conventional technologies and there is also a series of interesting applications that such a device can cover. Such applications may be: 1) monitoring of air quality in residential environments and space, 2) checking of foodstuff spoilage, 3) disease diagnosis, 4) even the replacement of the human nose eventually (15). With the research going on it is predicted that the price of an electronic nose will be $1 by 2020, given that most electronic noses are currently priced from $20.000 to $100.000 (16).

2.1.2 Metal oxide sensors: principal of operation

In Figure 3 one can observe a simple schematic of a MOS sensor, in order to comprehend its principle of operation easier. Specifically, the sensor is constituted by an electrode (most often an interdigitated one), imprinted on an insulating substrate (most often SiO_2), on top of which the sensing film is deposited (the thicknesses range from 10 to 300mm for thick and 6-1,000nm for thin films) and by an integrated heater: a resistive material that heats the sensing film through the Joule effect, after the application of appropriate voltage. The common range of operational temperatures is 200-500 °C.

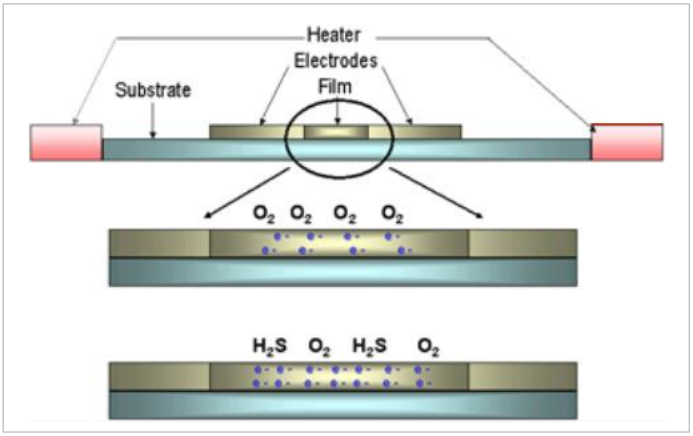


Figure 3: Schematic of MOS sensor and its basic reaction **(17)**

The sensing film has the property of changing its conductance after the chemical reaction with an appropriate gas sample in a way that is proportional to the concentration of the gas.

The MOS sensors belong to two classes: n-type (sensing film: zinc oxide, tin dioxide, titanium dioxide or iron (III) oxide) and p-type (sensing film: nickel oxide, cobalt oxide). The former corresponds to sensing films whose surface reacts with the oxygen in the air, resulting in trapping any free electrons of the surface or the grain boundaries of the sensing film. Since the number of the free carriers decreases, depleted regions are created – which means formation of potential barriers between the grains that prohibit the mobility of any carriers- and hence the resistance increases in the respective regions. In presence of reducing gases (H2, CH4, CO, C2H5 or H2S) that react with the trapped oxygen of the surface, the resistance

drops because the electrons are released back and are free to flow between the grain boundaries. On the other hand, the p-type class corresponds to sensors whose resistivity decreases after the reaction of the oxygen with the so called oxidizing gases (O2, NO2, and Cl2). This is because in that case the gases deplete the electron depleted regions furthermore and hence they produce holes that flow between the grain boundaries. The above mentioned reactions of the n-type sensors are presented by the following formulas (1) and (2), where R represents the reducing gas:

$$\frac{1}{2}O_2 + e^- \rightarrow O^- \quad (1)$$

$$R + O^- \rightarrow RO + e \quad (2)$$

At this point it becomes understood that the sensing film is basically a variable resistor, whose resistance is being measured through the electrodes on which that film has been deposited and its corresponding values are basically the electrical outputs of the sensor. The output signal that corresponds to the resistance of the film in the absence of any gas is called baseline (Rair), while the one that corresponds to the resistance of the film in the presence of the analyte (Rgas) is the response of the sensor. The sensitivity of oxide gas sensors may be expressed in several ways. The equations (3) and (4) are the ones that have been used to express the sensitivity of the sensors that have been measured in the frame of the present thesis and as it is shown below they are basically the two sides of the same coin (18).

$$S = \frac{\text{Rair} - \text{Rgas}}{\text{Rair}} \rightarrow (3)$$

$$S = 1 - \frac{\text{Rgas}}{\text{Rair}} = 1 - \frac{1}{S^*}, \qquad S^* = \frac{\text{Rair}}{\text{Rgas}} \quad (4)$$

The sensitivity is a function of the baseline value (Rair), which does not remain stable over a long period of time, resulting in problems of consistency and accuracy when it comes to labelling a known input (i.e given concentration of known gas) to a certain feature (i.e sensitivity). It is therefore this drift of the baseline that causes the so called instability of the sensors. Other factors such as: 1) film thickness 2) doping in the film, 3) grain size, 4) operational temperature of the film also need to be optimized when designing a sensor in order to achieve higher values of sensitivities (19).

2.1.3 Electronic nose: principles of operation

The electronic nose was first developed with the aim to mimic the human olfaction system. For this reason, the description of its principles of operation will be realized with references to the olfaction system in order to simplify the concepts when necessary. In Figure 4 such an analogy is illustrated and the reader can refer to it for purposes of comprehension.

The electronic nose consists of a sample delivery system, an array of sensors - which constitutes the mechanism for the chemical detection of an odor - and a mechanism for pattern recognition, such as a neural network (19, 20, 21). The array of sensors play the role of the receptors in the human nose, while the pattern recognition mechanism plays the role of the cognitive process that takes place in the human brain.

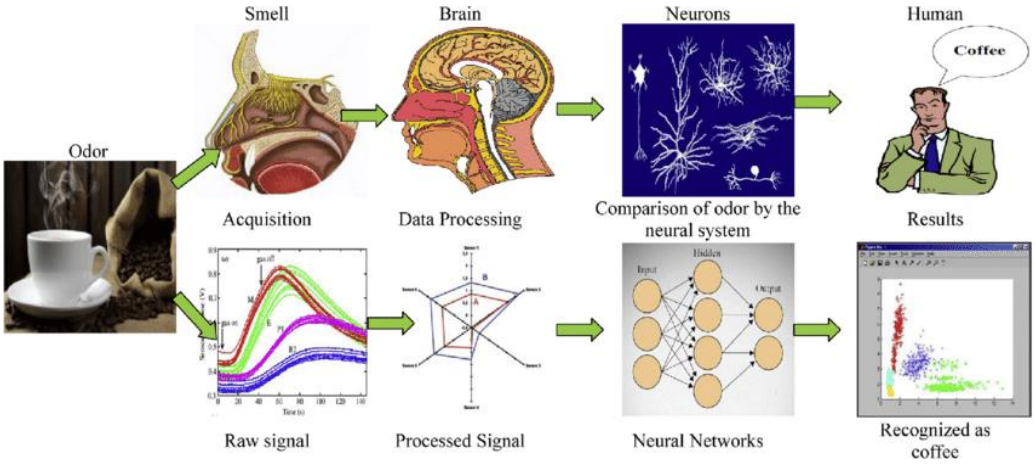


Figure 4: Analogy between the electronic nose and the human olfactory system (18)

In order for the human brain to conclude that the smell of coffee corresponds to coffee, it is mandatory that the smell is already known. In other words due to experience, the brain has already stored the special characteristics of the smell and has linked them to a specific object (i.e coffee). Every time that the same characteristics are experienced the brain is capable of recalling that object. It becomes clear that in the case of an artificial brain in order to match a pattern – i.e electrical output of a specific gas of certain concentration – it is needed that the pattern is known beforehand. This is achieved by "feeding" the artificial brain with the appropriate characteristics (those corresponding characteristics are called features in the field of machine learning) that the electronic nose is supposed to correspond to specific gases and concentrations. While the human brain is capable of determining the special characteristics of a smell by itself, in the case of the artificial brain the features are carefully extracted during the data analysis of the electrical outputs of the sensors. Such a feature might be for example the sensitivity that a certain sensor presents after it has chemically reacted with a certain gas of certain concentration.

After the brain has processed the signal of the smell/gas and has determined the features, it compares them to features of which the brain is already aware of. In the human brain this comparison is realized by the neurons. Similarly, in the case of the artificial brain the artificial intelligence that performs the classification of the gas is called neural network. By the end of this process, both brains provide with an answer. In the case of the artificial brain, the result will answer to the question: which gas and of what concentration did the sensors-array sense?

The part that concludes the description of the device is the delivery system. Depending on the application, there are different ways that can deliver the gas to the detection system of the electronic nose. A common system, where a volatile specimen is sampled consists of a tube

that collects the air with the help of a pump so that the air flows in the sensor chamber where the sensors array resides. In cases where the specimen is solid the system is modified to directly inject the specimen in the sensor chamber (22).

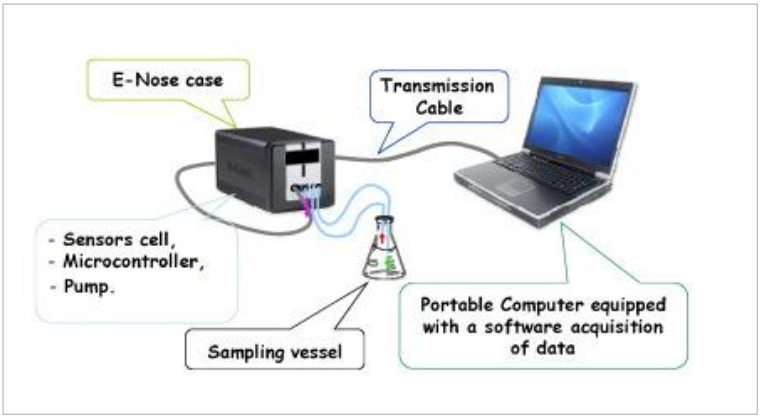


Figure 5: **Schematic representation of electronic nose (23)**

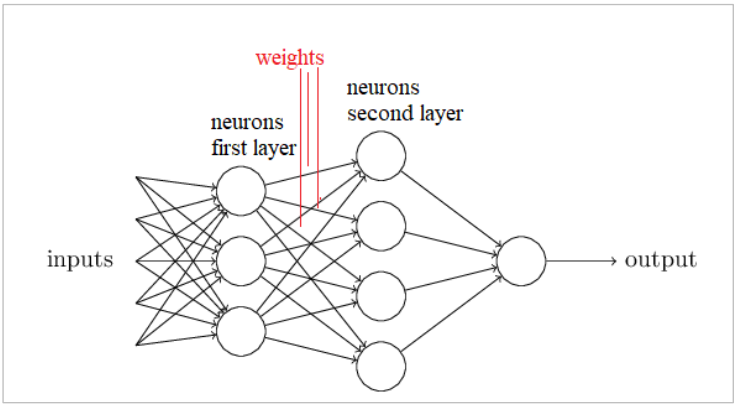
Figure 5 presents the simplified schematic of an electronic nose and the respective parts that have been analyzed in this chapter.

2.1.4 Data process and pattern recognition of complex systems

As it was explained above, the artificial neural network is the way to go in order for an electronic nose to be able to detect and identify experimental samples. In order for the electronic nose to "learn" how to classify the gases appropriately, a learning environment is established and a training algorithm is applied.

In Figure 6 one can see the representation of a neural network. Since this is a computational platform that mimics the human neural network, each node of the network is called neuron. The learning and training algorithm resemble the process during which the knowledge is acquired and which is supposed to be stored. The storage is realized through the connections between

the neurons. One can imagine that the knowledge is stored through the communication of one neuron with its successive neuron and so on and so forth. It can be understood that this information which is exchanged between successive neurons is of major importance, and it is basically the amplitude of this connection, which is known as synaptic weight in the field of machine learning (24).



The known data – or else labeled data – are fed in the network as inputs. The labeled data get classified by the network but since their correct output is known the network can adjust its classification hypothesis by evaluating the error of its prior classification attempt

Figure 6: Architecture of a two-layer neural network **(25)**

classification attempt. The network will use the same hypothesis for unknown data and hence it will be able to classify them correctly. In more detail, it is the modification of the weights that takes place by the application of the labeled data and their final values which minimize the difference between the desired output and the real output of the network. Ultimately those values formulate the classification hypothesis. The property of the neural networks to classify correctly unknown data is called generalization (26).

In Figure 7 the mathematical model described above is illustrated. The values xl, x2… xn represent the input examples (in our case gases of specific concentrations accompanied with their related features). Each xi is then multiplied with the respective weight wi of each neuron (wi resembles the synapse of the neuron in the human brain). The next step is the summation of the respective xi*wi multiplications (summation junction), which later on gets "filtered" by a transformer (that transformer is called activation function and in most cases it is a sigmoid function) in order to produce the final output. The final output gives basically the classification result and corresponds to a finite value (27). Till this point the input signal of the network has been propagating forwards (from the input till the final output). Since the signal reaches the output, the error is generated and starts to propagate backward in the network, modifying the weights. This process repeats itself up to the point that the error is minimized. The minimization of the error is achieved most commonly by the so-called gradient descent method (28).

At this point it makes sense to also mention that in most cases the input data of the neural network are data that have been undergone a process called PCA (Principal Component Analysis).

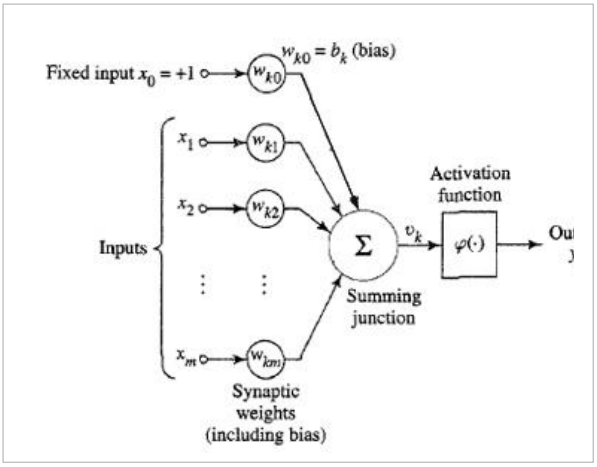


Figure 7: Architecture of a two-layer neural network **(25)**

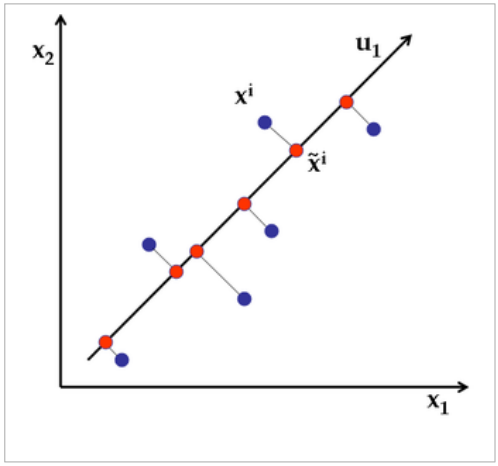


Figure 8: PCA: 2d to 1d transformation **(30)**

PCA is a statistical procedure that uses an orthogonal transformation in order to convert possibly correlated variables into linearly uncorrelated variables

In the field of machine learning, each feature represents a dimension and hence one can end up dealing with a multi-dimensional hyperspace. Even if the features are not correlated, PCA

can be then applied for reasons of visualization, which means that PCA can be applied to reduce the original space up to a three dimensional space.

In Figure 8 the general concept of PCA is shown. It can be shown that by solving the eigenvalue problem, the directions of maximum variance of the original data is found. Those directions correspond to vectors on which the same data are then projected and hence minimizing the space, ending up with new data that are the linear combinations of the original ones (29).

In the same Figure 8, x1 and x2 correspond to two different features, the blue dots correspond to the original data, u1 is the vector of the maximum variance of the data on which the blue dots are projected and their projections correspond to the orange dots. The orange dots are the new data and are plotted on a one dimensional space.

2.2 BIBLIOGRAPHY

(1). Persaud K., D. G. (1982). Analysis of discrimination mechanisms in the mammalian olfactory system using model nose. Nature, 299, 352

(2). Halova, J., Strouf, O., Zak, P., Sochozova, A., Uchida, N., Yuzuri, T., Sakakibara, K., Hirota, M. Quant. (1998). Potential Application of Electronic Olfaction Systems in Feedstuffs Analysis and Animal Nutrition. Struct.-Act. Relat. 17, 374

(3). Gaillard, I., Rouquier, S., Giorgi. (2004). D. Olfactory Receptors. Cell. Mol. Life Sci, 61, 456

(4). Shirokova, E., Schmiedeberg, K., Bedner, P., Niessen, H., Willecke, K., Raguse, J. D.; Meyerhof, W.; Krautwurst, (2005). Identification of specific ligands for orphan olfactory receptors. G protein-dependent agonism and antagonism of odorants., D. J. Biol. Chem. 280, 11807

(5). Vidic, J. M., Grosclaude, J., Persuy, M. A., Aioun, J., Salesse, R., (2006). Quantitative assessment of olfactory receptors activity in immobilized nanosomes: a novel concept for bioelectronic nose, Pajot-Augy, E. Lab Chip 6, 1026.

(6). Liu, Q. J., Cai, H., Xu, Y., Li, Y., Li, R., Wang, P. (2006) Olfactory cell-based biosensor: a first step towards a neurochip of bioelectronic nose. Biosens., Bioelectron., 22, 318.

(7). Schoning, M. J.; Poghossian, A. (2006) Bio FEDs (Field-Effect Devices): State-of-the-Art and New Directions Electroanalysis, 18, 1893.

(8). Ballabio, D., Cosio M. S., Mannino S., Todeschini, R. (2006) A chemometric approach based on a novel similarity/diversity measure for the characterization and selection of electronic nose sensors. Anal. Chim.Acta, 578, 170.

(9). Gardner, J. W. Boilot, P., Hines, E. L. (2005) Enhanced electronic nose performance by sensor selection using a new integer based genetic algorithm approach Sens. Actuators, B,106, 114.

(10). Hajek, P.; Havranek, T. (1978). Mechanizing hypothesis formation mathematical foundations for general theory; Springer-Verlag: Berlin, Heidelberg, New York.

(11). Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J. (2004) Direct Determination of Urinary Creatinine by Reactive-Thermal Desorption-Extractive Electrospray-Ion Mobility-Tandem Mass Spectrometry *Analyst*, *129*, 984.

(12). Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L. (1988) Chemometrics: A Textbook; Elsevier: Amsterdam.

(13). Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., Smeyers-Verbeke, J. (1998) Handbook of Chemometrics and Qualimetrics; Elsevier: Amsterdam.

(14). Margaret Amy Ryan, Hanying Zhou, Martin G. Buehler, Kenneth S. Manatt, Victoria S. Mowrey, Shannon P. Jackson,Adam K. Kisor, Abhijit V. Shevade, and Margie L. Homer (2004) "Monitoring Space Shuttle air quality using the Jet Propulsion Laboratory electronic nose", IEEE sensors, vol. 4,n 3

(15). Frank Rock, Nicolae Barsan, and Udo Weimar. (2007) Electronic Nose: Current Status and Future Trends, Morgenstelle, 15, 72076

(16). Shih-Wen Chiu and Kea-Tiong Tang (2013)"Towards a Chemiresistive Sensor-Integrated Electronic Nose: A Review", MDPI Sensors, 14214–14247

(17). K. Arshak, E. Moore, G.M. Lyons, J. Harris and S. Clifford. (2004) A review of gas sensors employed in electronic nose applications, Emerald Group Publishing Limited · ISSN 24, 181–198

(18). Retrieved from http://www.enose.nl/rd/technology/

(19). Z.Haddi. A.Amari. B.Bouchikhi. (2011) A portable electronic nose system for the identification of cannabis-based drugs, Sensors and Actuators B: Chemical, 155, 2

(20). A. Sayeed and M. S. Shameem. (2011) Electronic Nose, Advances in Medical Informatics, vol 1

(21). D. G. Persaud K. (1982) A chemical-detecting system based on a cross-reactive optical sensor array Nature. 1982

(22). J. Gardner and P. Bartlett. (1994) A brief-history of electronic noses, Sens.Actuators B Chem, 18, 211–220

(23). D. Karamousantas, G. Chatzarakis, D. Oikonomou, L. Ekonomou, and P. Karampela. (2008) Effective insulator maintenance scheduling using artificial neural networks, IET Generation, Transmission and Distribution, 382, 697 - 700

(24). S. Haykin. (1999) Neural Networks: A Comprehensive Foundation, Second Edition

(25). Derived from https://www.doc.ic.ac.uk/project/2015/163/g1516306/index.html.

(26). G. D. (2007) Principles of artificial neural networks, World Scientific Publishing, 2nd Edition

(27). D. Bishop. (1995) Neural Networks for Pattern Recognition. Clarendon Press, Oxford

(28). K. Hornik, M. Stinchcombe, and W. Halbert. (1989) Multilayer feedforward networks are universal approximators, Neural Networks, 2

(29). P. M.Jainl and V.K.Shandliya. (2013) A survey paper on comparative study between principal component analysis (pca) and exploratory factor analysis (efa). International Journal Of Computer Science And Applications, 328, 379 -384

(30). https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.PCA

3.1 EXPERIMENTAL PART

3.1.1 Description of the experimental setup

In this part of the thesis the experimental setup will be described, the core of which are the instruments presented in Figure 9. The Owlstone instrument is depicted in Figure 9a, which hosts one single oven for an acetone permeation tube and the VICI instrument is depicted in Figure 9b which hosts two ovens, for a toluene and a benzene permeation tube respectively. The former's temperature is set at 40 oC and the latter's temperatures are set at 60 oC and 50 oC, while the respective boiling temperatures are: 56 oC for acetone, 111 oC for toluene and 80 oC for benzene.

a. b.

Figure 9: a. Owlstone permeation instrument b. ViCi permeation instrument

The tubes in the oven are being heated at a constant temperature, which is always lower than the boiling temperature of the liquid material contained in the tubes. In that way, the material in the tube is in a dynamical phase of equilibrium between its gas and liquid phase. Over time, the liquid evaporates and permeates through the walls of the tube at a constant rate (R), which depends on the temperature of the oven

With the help of those two instruments: acetone, toluene and benzene are driven along a line, where the gases can be furthermore diluted by another analyte before they reach the chamber in which a sensor board is located. Each instrument is characterized by a different internal set-up, and hence the final concentration of the gas inside the chamber is calculated with different equations. In Figure 10, one can appreciate the schematic of the internal set-up of each instrument.

As it can be observed in Figure 10a, Owlstone has an internal exhaust system, which means that the user defines the total flow of the acetone (F) and also the flow (out of the total flow), which is going to be led to the exhaust (Fexhaust). The sensor board will sense therefore the remaining flow (Fin) which gets diluted by dry air (Fdilution2, corresponds to Dry air 2 in the schematic) before it reaches the chamber. This description can be mathematically expressed in terms of flows and concentrations (C), as in equation 5.

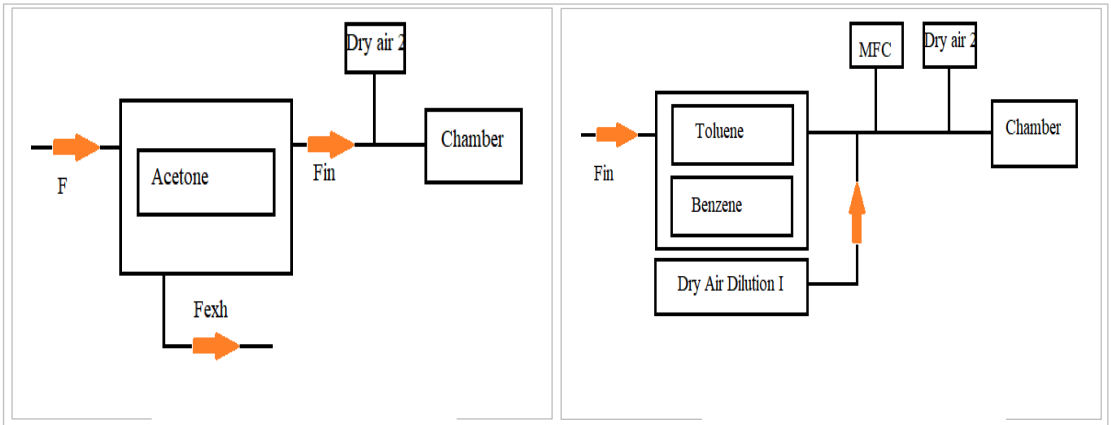


Figure 10: a.Internal schematic of Owlstone b.Internal schematic of Owlstone ViCi

$$C = \frac{R}{F_{in}+F_{exhaust}} * \frac{F_{in}}{F_{in}+F_{dilution2}} \quad (5)$$

In comparison to the Owlstone, the VICI gives to the user the ability to determine the flow of either toluene or benzene but there is no internal exhaust. On top of that, there is an internal dry air dilution (Fdilution), whose flow can be determined by the user. Because of the fact that this specific set-up does not provide a wide range of different concentrations for the purposes of the experiment, the set-up was upgraded by adding an external mass flow controller. The mass flow controller practically plays the role of an exhaust. In other words, while the user determines the Fin+Fexhaust at the input of the Owlstone instrument, in the case of the VICI she just determines the Fin at the input of the upgraded VICI instrument. The mathematical equation

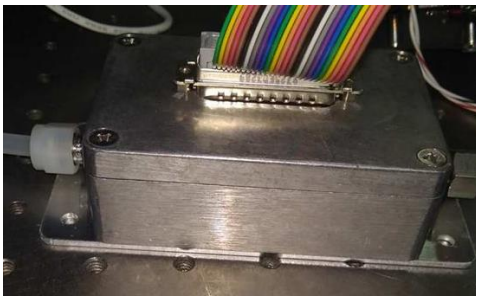
6 describes the effect of the set-up on the total concentration of gas that the sensor board ultimately senses.

$$C = \frac{R}{Fin} * \frac{F_{in} - F_{exhaust}}{F_{in} - F_{exhaust} + F_{dilution} + F_{dilution2}} \quad (6)$$

The two instruments cross-meet before the chamber and user-controlled valves are used to let in either the gas flow coming from the Owlostone or the gas flow coming from the VICI. In Figure 11a one can observe the controller with which the respective valves of the dilution and of the instruments can be set at either on (flow goes in) or off mode (flow goes to exhaust) and in Figure 11b the chamber is shown.

In the chamber box there is the possibility to place boards on which six sensors can be mounted and hence measure the response of all six sensors simultaneously for a certain concentration of gas flowing into the chamber. Such boards can be shown in Figure 12.

a.



b.

Figure 11: a. Valve controller b. Chamber of sensor boards

Figure 12: Three different sensor boards that normally get hosted by the chamber

In order to evaluate the response of each sensor at the presence of a certain gas, it is mandatory to monitor the changes of their respective resistances. On top of that, as it has been mentioned in the previous chapter, it is mandatory for each sensor to be heated. In Figure 13, the schematic of the electronic circuit for the read-out and the heating is shown.

The six sensors, and specifically their respective sensing films, play the role of a resistance, the value of which is approximately given by the manufacturer for sensors operating under specific set of conditions (operating temperature, humidity, ambient temperature, presence of

gases). It should be noted that the uncertainty on the specified resistance values for a given sensor is always very significant, sometimes exceeding 50%. Therefore, each sensor requires individual evaluation to assess its properties. Rsensor_1, Rsensor_2… Rsensor_6 do represent the six sensing films of the sensors. For each sensor one load resistance in connected in series: Rload_1, Rload_2… Rload_6, while a voltage drop of 1.8V is applied across every pair of sensor resistance-load resistance. The value of the load resistance is known by the user. Specifically, the load resistor is an adjustable resistor, the value of which is adjusted by the user in order to decrease the noise of the measurement. In Figure 14, one can see the six adjustable resistors mounted on an external board which is connected in series to the chamber.

From the circuit of Figure 13, six outputs are obtained and their read out is graphically illustrated on the screen of a computer at real time. The voltage output of each single sensor resistance-load resistance pair is being measured across the load, as it can be seen in Figure 15. The Figure 15 is basically a typical voltage divider circuit, for which the output voltage is given by the equation 7.

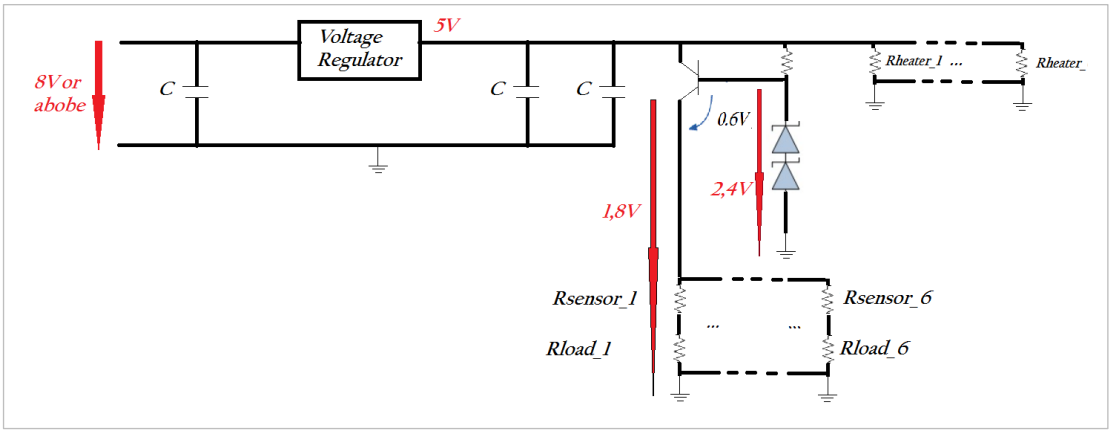


Figure 13: Schematic of the electronic circuit for the read-out and the heating of the sensor board

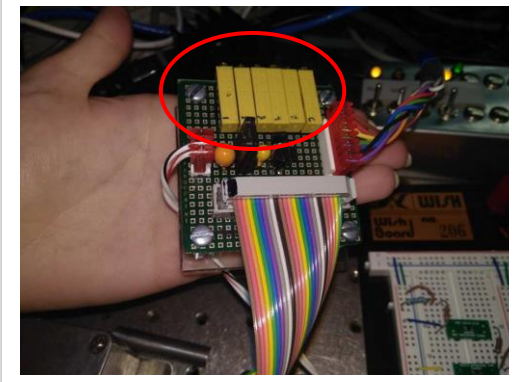


Figure 14: The six adjustable load resistors of the circuit are in the red circle

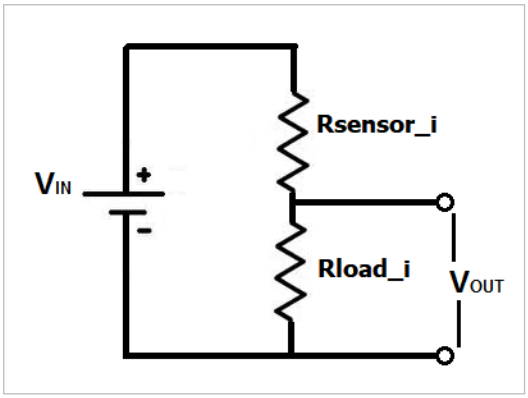


Figure 15: Output voltage for every sensor resistance-load resistance pair

$$Vout = Vin\frac{Rload\_i}{Rload\__i + Rsensor\_i}\ , \mathrm{i} = 1,2,\ldots 6 \quad (7)$$

The equation 7 has been used to convert the obtained voltage data into resistance values, since the resistance values are widely used in the bibliography when evaluating sensitivities and the performance of sensors. Since the Vout is the measured value and both Vin and Rload_i are known parameters, the resistance of the sensing film can be calculated. In order to establish automatic voltage – resistance conversions, a script in python has been written by the author of the present thesis. The relevant scripts can be found in Appendix A.

It is worth mentioning, that the value of the adjustable resistance is of great importance while conducting the measurement, which can be understood while paying attention at the circuit of Figure 15. For a given Rsensor_i value, the higher the Rload_i the higher will be the drop on the load and hence the output voltage. However, the Vin is restricted to 1.8V by the supply voltage of the circuit. That means that in case the adjustable resistance is adjusted in such a way that the output is close to 1.8V in dry air conditions, when the gas (acetone/benzene/toluene) is purged in the chamber, the resistance of the sensor will decrease (the voltage drop across the Rsensor_i will decrease), and hence the drop voltage across the load will increase even more. The latter increase of the output voltage however is restricted as already said and that suggests that there is risk of degrading the read-out resolution.

In Figure 16 the power supply of the experimental set-up is shown. As long as the supply provides a voltage of 8V or above, a voltage regulator which is depicted in the circuit schematic of Figure 13, gives always an output of 5V in order to protect the circuit, while the extra capacitors before and after the voltage regulator provide stability to the circuit. Those 5V are applied to the six integrated heater resistors of each sensor, which are connected in parallel. It is because of this voltage drop across the Rheater_i that the sensing film of each sensor get heated through the Joule effect.

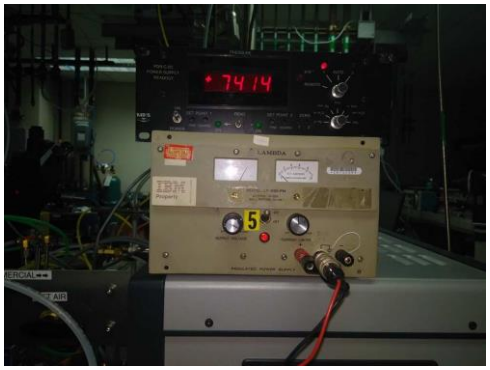


Figure 16: Power supply of the experimental setup

In parallel to the heaters, two Zener diodes and a resistance are connected in such a way that the drop voltage across the resistance is 2.6V (= 5V – 2.4V based on the schematic of Figure 13) and is applied at the input of a BJT transistor. Since, the intrinsic drop of voltage between the base and the emitter is always 0.6V and the voltage drop across the Zener is 2.4V, based on the schematic of Figure 13, a voltage of 1.8V (2.4V – 0.6V = 1..8V) is always applied across each Rsensor_i – Rload_i pair, as it has already been mentioned.

At this point it should be pointed that it is not the operational temperature that it is typically disclosed by the manufacturer but the voltage to be applied at the sensor's heater. This is particularly relevant in the context of one experiment that took place in the framework of this thesis that will be introduced in the following chapters.

3.1.2 Description of the experiments

The measurements from which the data where obtained for the first part of the thesis, were conducted by the author and with the help of the experimental set up described in the previous chapter. Two boards were tested on which six different sensors were mounted. Subsequently, twelve different sensors have undergone different experimental procedures in total. Table 1 contains the names of all the sensors per board.

Those 12 sensors were exposed to the three gases – acetone, toluene and benzene – of different concentrations. The concentrations were calculated by the author of this thesis with the help of the equations 5, 6. The permeation rate (R), which is included in the equations 5 and 6, was calculated for each gas in the following way: each tube was filled with the suitable analyte and its weight was measured before inserting it in the oven. For a period of time the tubes were taken out of the oven in order to record the new weights. The rate of the weight loss was calculated from the slope of the weight data points over time, after applying a linear fit and it was then converted into a rate of mol loss over time, by using the molecular weight of the analyte. In Figure 17 an example is given for such a conversion for the case of benzene.

Table 1

	BOARD I	BOARD II
Sensor 1	TGS2602	1330
Sensor 2	TGS2611	2330
Sensor 3	TGS2620	10330
Sensor 4	MQ138	ASMCVP2
Sensor 5	MQ137	TGS8100
Sensor 6	TGS2600	5914

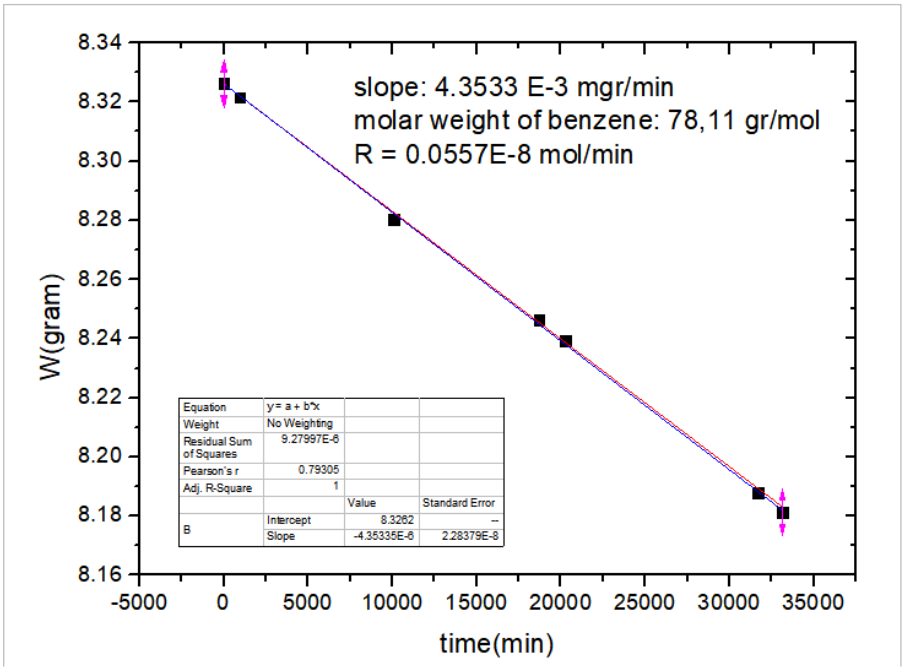


Figure 17: Experimental weight data points and making use of its known molecular weight

One series of experiments consisted in recording the voltages (which were afterwards converted in resistances: Appendix A) for different concentrations of each gas (acetone, toluene and benzene). The concentrations of interest were the following: 150, 200, 500, 1000, 1250, 1500, 2000, 3100, 3600, 4380, 4380 and 5700 ppb. The aim of this experiment was the extraction of sensitivity vs concentration graphs. Such graphs provide valuable information, in order to evaluate whether the same sensor gives a different response (in terms of sensitivity) for the same concentration of each gas. In other words, if one sensor responds differently for each gas then for an unknown input (input refers to some gas of some concentration) it is possible to make a conclusion about the nature of the gas (determine whether it is acetone, benzene or toluene in our case) and its respective concentration. Practically, since the response of the sensors is pretty similar, the ultimate goal of such an experiment is to determine which four sensors out of the twelve provide the most distinctive responses when compared to each other so as to combine them on the same board in order to make such conclusions about the nature and the concentration of the purging gas. This series is named: "Sensitivity series" and the measurements were realized by the following this pattern: 1. purging dry air in the chamber, 2. purging gas of concentration x1 in the chamber, 3. purging dry air in the chamber 4. purging gas of concentration x2 in the chamber, 5. purging dry air in the chamber and so on and so forth. The x1, x2 values represent the concentration values of interest that have already been mentioned, while the gas is either acetone or benzene or toluene.

However, since the instability of the sensors is a big issue (see chapter 1) two other series of experiments were realized in order to reassure that the obtained results of their response is fairly stable over the course of a period of time. For this reason, a second series of experiments consisted in recording the response of the sensors for the same concentration of a gas repeatedly, over a short period of time (hours). This series is named: "Reproducibility series", and all the responses were recorded one after the other following this pattern: 1. purging dry air in the chamber: 2. purging gas of concentration x1 in the chamber, 3. purging dry air in the chamber, 4. purging gas of concentration x1 in the chamber and so on and so forth.

A third series called: "Sensitivity over time series" was carried out and is similar to the "Reproducibility series" with the difference that the responses were recorder after long delays (usually several days) instead of after a couple of hours . The pattern of this experiment can be described as following: Day #1: 1. purging dry air in the chamber, 2. purging gas of concentration x1 in the chamber. Day #2: 1. purging dry air in the chamber, 2. purging gas of concentration x1 in the chamber. Day #3: 1. purging dry air in the chamber, 2. purging gas of concentration x1 in the chamber and so on and so forth. The "Reproducibility series" are ideal in order to observe whether the same sensor gives different responses in a very short period of time, where the environment is theoretically maintained constant. Any differences in the response in such short period of time are not desirable and hence the sensor is considered to be not ideal and hence not a suitable candidate for the final selection of the four-sensors board that has been mentioned in this chapter. On top of that, the experimentalist may observe certain fluctuations which are more related to the system of the experimental set-up (f.e slight

humidity, temperature, pressure variations) than to the sensor itself. The "Sensitivity over time series" gives a more general idea of the performance of the sensor, since the experimentalist can keep track of the sensitivity of the sensor and how it might be affected by the drift of the baseline (corresponds to the Rsensor_i when only dry air is purging and no gas is present) over time. An ideal sensor is expected to have minimal drift of its baseline, which is the figure of merit when talking about stability issues.

At this point a further clarification should be made since it might seem ambiguous the fact that a commercial sensor, which is already on the market, might show bad performance in the "Reproducibility series". The first reason behind this fact is that the conditions under which the sensors are operating are not the ones that are recommended by the manufacturers. Specifically, all datasheets accompanying commercial sensors give sensitivity values under humidity conditions. However, the experiments that have been carried out during the present thesis were focused on dry air conditions (low relative humidity) only, as it was thought that this would allow for better control of the working environment and, consequently, enable to gain a clearer understanding of the sensors behavior. The second reason is that the majority of the sensors are described as sensors for: cigarette smoke, cooking odors, gaseous air contaminants or VOC detection. In other words, there is no specific reference for detection of certain gases, which means that acetone, toluene and benzene may either not be suitable targets for specific sensors or may be even poisonous (i.e., causing irreversible adsorption at the sensor surface). The third reason is that the average concentration for which the sensitivity is calculated in the datasheets of the commercial sensors is 1000ppm (=1000.000ppb), while the concentration for which the sensitivity was calculated in "Reproducibility series" is just 1000ppb because this level of concentration is more relevant for most of the applications targeted by the team.

All the above mentioned series have in common the fact that during the experiments the temperature (the voltage of the power supply) has been always stable. This mode of operation is named: "Single Temperature Operation (STO)". Another mode of operation has also been applied for one single sensor, where the temperature of the heater is not the same over the course of an experiment but the concentration of the purging gas is the same. The name of this mode is: "Temperature Controlled Operation (TCO)". The author of the present thesis did not take part in collecting TCO data but carried out the data analysis and developed the machine learning model. For this reason, only the working principles of the experiment will be described -and not the experimental details in an inclusive way – in order to lay down the foundations for the machine learning part that will be presented in the next chapter.

In Figure 18 one can appreciate the principal of the TCO. Since, as it has been mentioned already, the temperature of the heater is not known, but the optimal voltage applied to the heater is known by the manufacturer, the changes of the temperature correspond to voltage changes. Specifically, in Figure 18 the term 100% corresponds to the value of voltage that is recommended by the manufacturer, the term 50% corresponds to a value of voltage that is 50% lower with respect to the recommended voltage etc. The absolute value of the operational temperature might not be known, but is not of significant interest either.

During the TCO the sensor is heated at a certain temperature for a couple of minutes in order to reassure that the system has achieved a steady state of operation. On a later step the temperature drops and it remains at this new temperature for the same period of time to restore a new steady state. This process is repeated multiple times

While the voltage (temperature) of the sensor is following the pattern depicted in Figure 18, dry air is purging in the chamber and the baseline is being recorded (Rair). After the completion of the run, the sensor follows once more the very same pattern of voltage, while a gas of certain concentration is purging in the chamber (Rgas). The ratio of Rgas/Rair is a figure of sensitivity (S*, see equation 4) for the sensor.

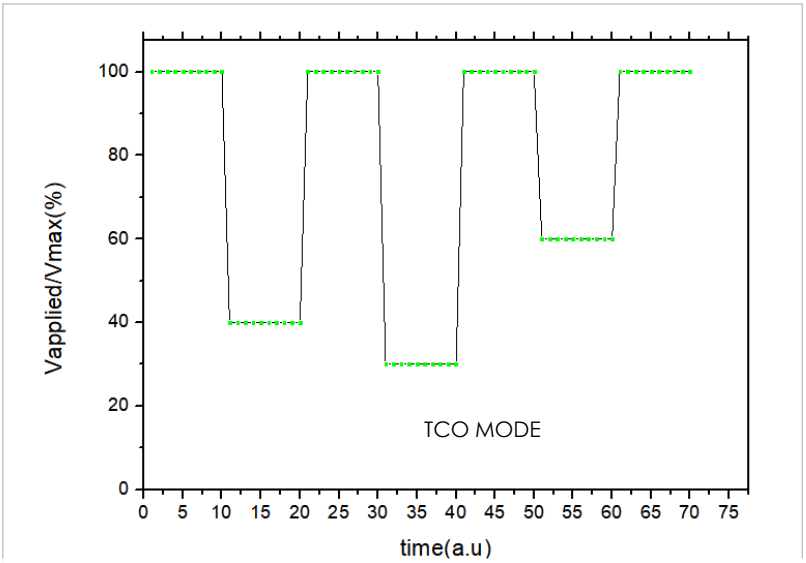


Figure 18: Example of the principal of operation of TCO

Similar measurements have been realized by other researchers and it has been shown in the recent bibliography of the field that the sensors tend to present specific response patterns every time the temperature drops and before it reaches again a new thermal steady state, even though the gas has been present all the time (I) in the environment. An example of this description is presented in Figure 19. It can be understood that from the perspective of machine learning, such phenomenon gives the opportunity to obtain more features for one single sensor and increase the accuracy of the prediction for what concerns the nature and the concentration of the gas when dealing with unknown inputs.

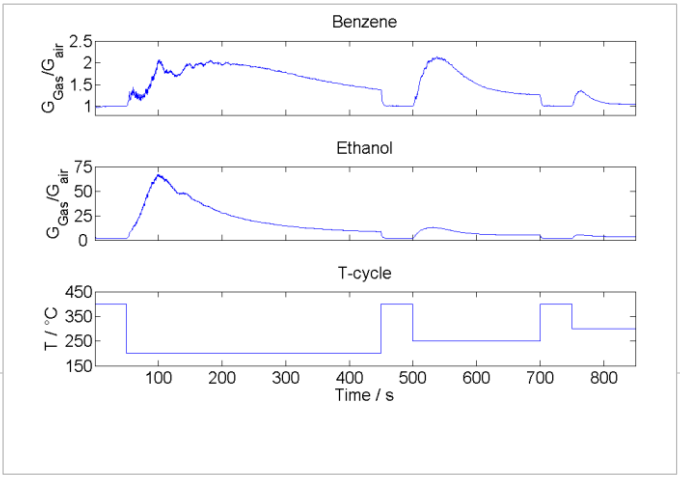


Figure 19: Example of a sensor's response when heated at different temperatures. The last graph represents the change in temperature of the heater, while the first two graphs represent its response at the presence of benzene and ethanol. As it can be observed, the sensitivity of the sensor increases dramatically after the temperature drops and the response tends to saturate back to its initial value when approaching a new steady state.

3.1.3 Experimental results of the commercial sensors

In this chapter the experimental data of the commercial sensors will be presented, which were obtained in the way that it was described in the chapter above. It is reminded that the following experimental data correspond to STO mode. In Figure 20, one can observe the data that correspond to the "Sensitivity series", after they have been converted in values of Rsensor_i, using the script of Appendix A. The data that have been selected to be shown among the existing ones correspond to the gas toluene for board I and acetone for board II. However, Table 2 summarizes all the measurements that have been conducted concerning board I and board II.

Table 2

	BOARD I			BOARD II		
Sensitivity series	Acetone 250 - 1500ppb	Toluene 150 - 2000ppb	Benzene 250 - 5700ppb	Acetone 250 - 1500ppb	Toluene 150 - 2000ppb	Benzene 250 - 1500ppb
Reproducibility series	Acetone 250ppb 1000ppb 2000ppb	Toluene 250ppb 1000ppb 2000ppb	Benzene 250 ppb 1000ppb 2000ppb	Acetone 1000ppb	Toluene 1000ppb	Benzene 1000ppb
Sensitivity over time series	Acetone 1000ppb	Toluene 1000ppb	Benzene 1000ppb	Acetone 1000ppb	Toluene 1000ppb	Benzene 1000ppb

One can notice that the data of Figure 20a and 20b are divided in different regions by green lines. Each region corresponds to a dry-air purging followed by a gas purging of some specific concentration. By zooming in every region and extracting the Rgas and Rair values, the sensitivity can be calculated for every region (i.e for every given concentration of gas). In Figure 21 the reader can appreciate a region that has have been selected and the corresponding values of Rgas and Rair. The sensitivities have been calculated using the equation 3.

After extracting the sensitivities for all sensors, gases and concentrations the data were put together in order to obtain meaningful plots in order to extract conclusions about the sensors' performance. The sensitivity has been similarly extracted from the data obtained from the remaining experimental series: "Reproducibility series" and "Sensitivity over time series".

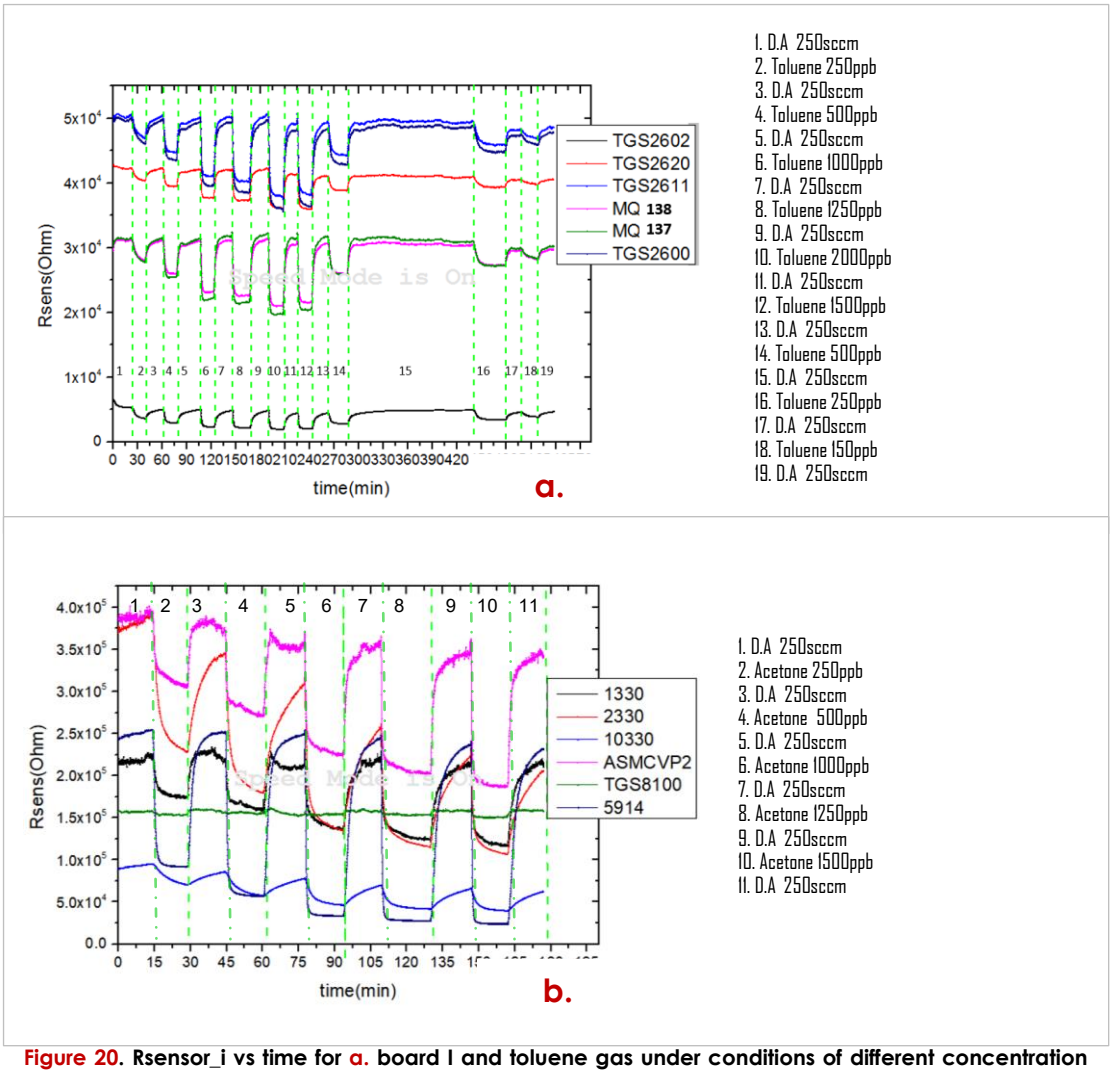


Figure 20. Rsensor_i vs time for a. board I and toluene gas under conditions of different concentration and b. board II and acetone gas under conditions of different concentration

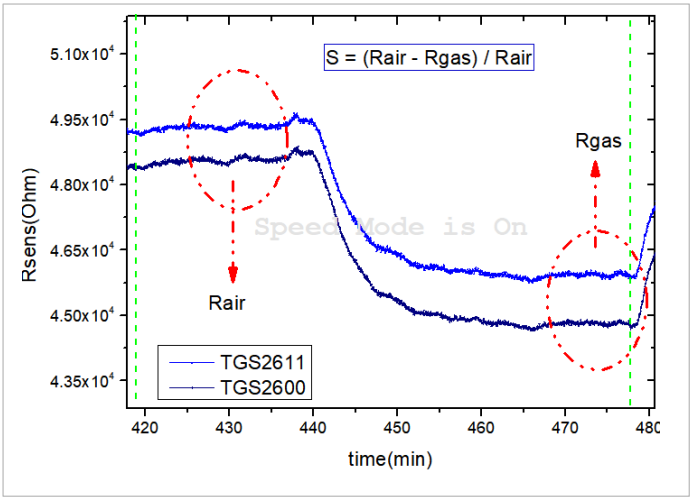


Figure 21. Zoom-in in region 16 of Figure 20a for sensors TGS2611 and TGS2600. The corresponding values of Rgas and Rair are pointed out, as well the equation with which the respective sensitivities have been calculated.

In Figure 22, the results of the "Sensitivity over time" are shown for the following cases: sensor TGS2602 of board I, sensor TGS2600 of board I, TGS2600 of board I, sensor MQ137 of board I and sensor 5914 of board II. The "y" axis of the plots presents the values of sensitivity corresponding to the response of 1000ppb of acetone, toluene and benzene, while the "x" axis

corresponds to the day on which the specific measurement has been performed in order to evaluate how the values of sensitivity evolve in time. It has to be noted that the first and forth sensors seem to have an exceptional behavior in terms of stability. The second and the third sensors seem to be less stable, especially because of some discrepancy in the case of toluene and benzene for day 1 and day 2. However, due to the fact that this discrepancy has exactly the same trend for both MQ137 and TGS2600, it is assumed that it is mostly attributed to systematic fluctuations. This assumption is supported by the fact that both sensors belong to the same board and hence the measurement has been taken simultaneity and whatever the fluctuations of the environment should affect all sensors. The TGS2602 seems to be unaffected by the same fluctuations but it also seems to be the most stable of all twelve sensors, which could justify its robust performance. The four sensors that have been selected to be presented are the ones with the best stability among all twelve sensors that have been measured. It can be understood that these four sensors serve as good candidates for the desirable four sensor board so far, in terms of stability.

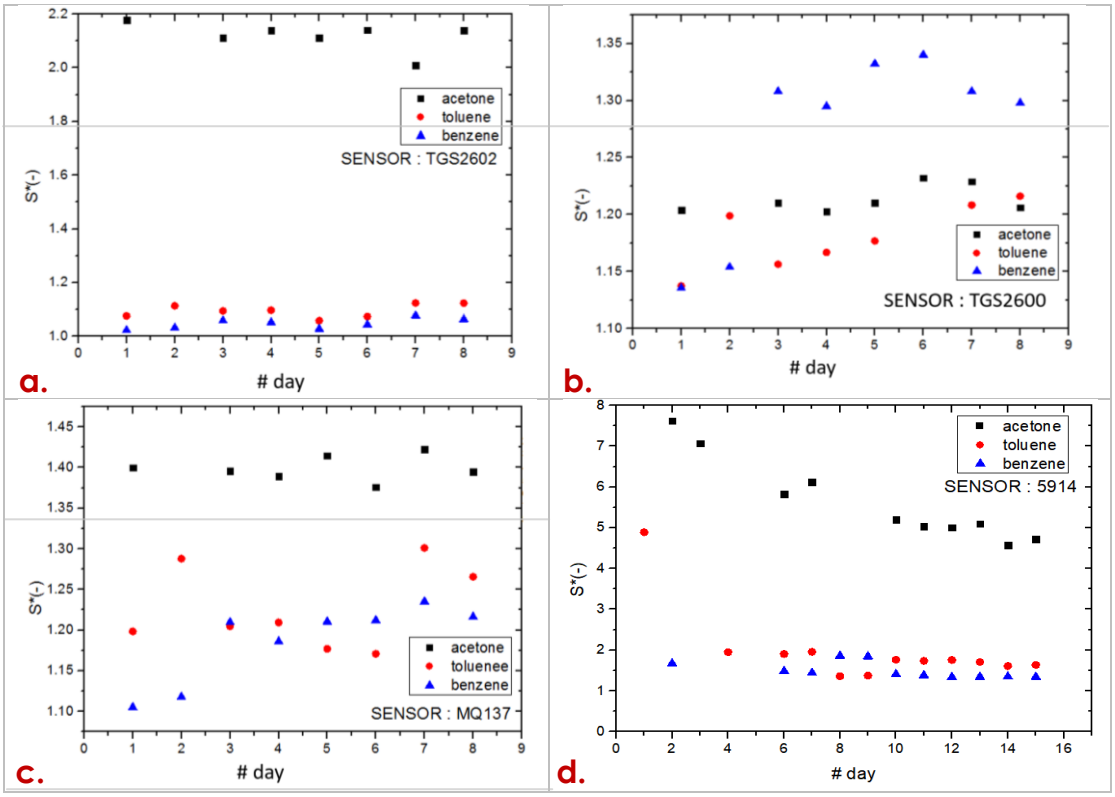


Figure 22. Sensitivity vs #day for a. TGS2602, board I b. TGS2600, board I c. MQ137, board I d. 5914, board II

However, it is not only the stability that is of great importance when it comes to sensor selection. As it has already been mentioned it is also important to evaluate whether the sensors show discriminant sensitivities for the same concentration of different gases. For this purpose the sensitivity over concentration for all gases are presented in Figure 23. The graphs that have been selected to be shown are the ones that correspond to the same sensors previously shown on Figure 22. Based on Figure 23 it becomes obvious the sensor TGS2602 and 5914 show sensitivities which are discriminant for same concentration of different gases for the case of

higher concentrations. Nevertheless, sensor 5914 presents a smaller but still detectable discrimination even for lower concentrations. Sensors MQ137 and TGS2600 show some discrepancy for the case of benzene, which might be attributed to the same systematic fluctuations. It is interesting to notice that these last two sensors can discriminate between acetone and toluene even better than TGS2602, at lower concentrations. This is a case, where if all those sensors are put on the same board one can complement the other (: TGS2602 and 5914 offer information about benzene, while TGS2600 and MQ137 discriminate toluene and acetone of low concentrations since the former sensors fail to do so).

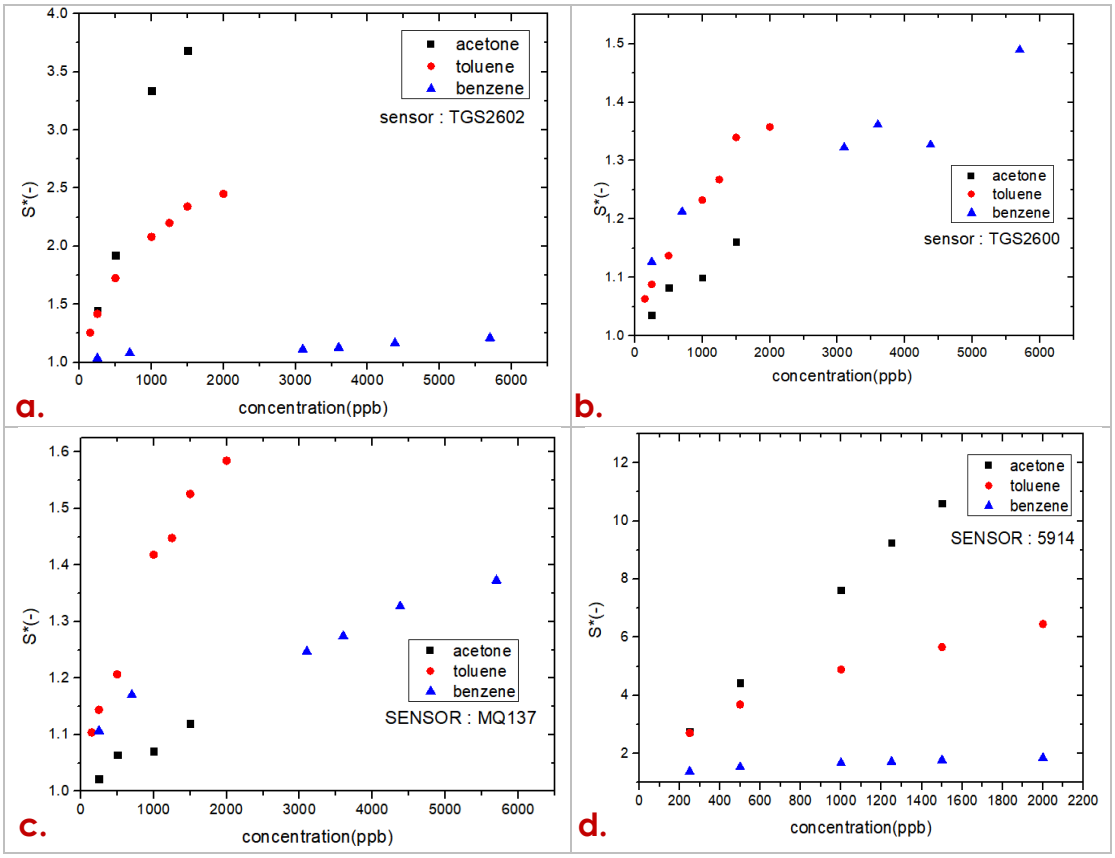


Figure 23. Sensitivity vs concentration for a. TGS2602, board I b. TGS2600, board I c. MQ137, board I d. 5914, board II

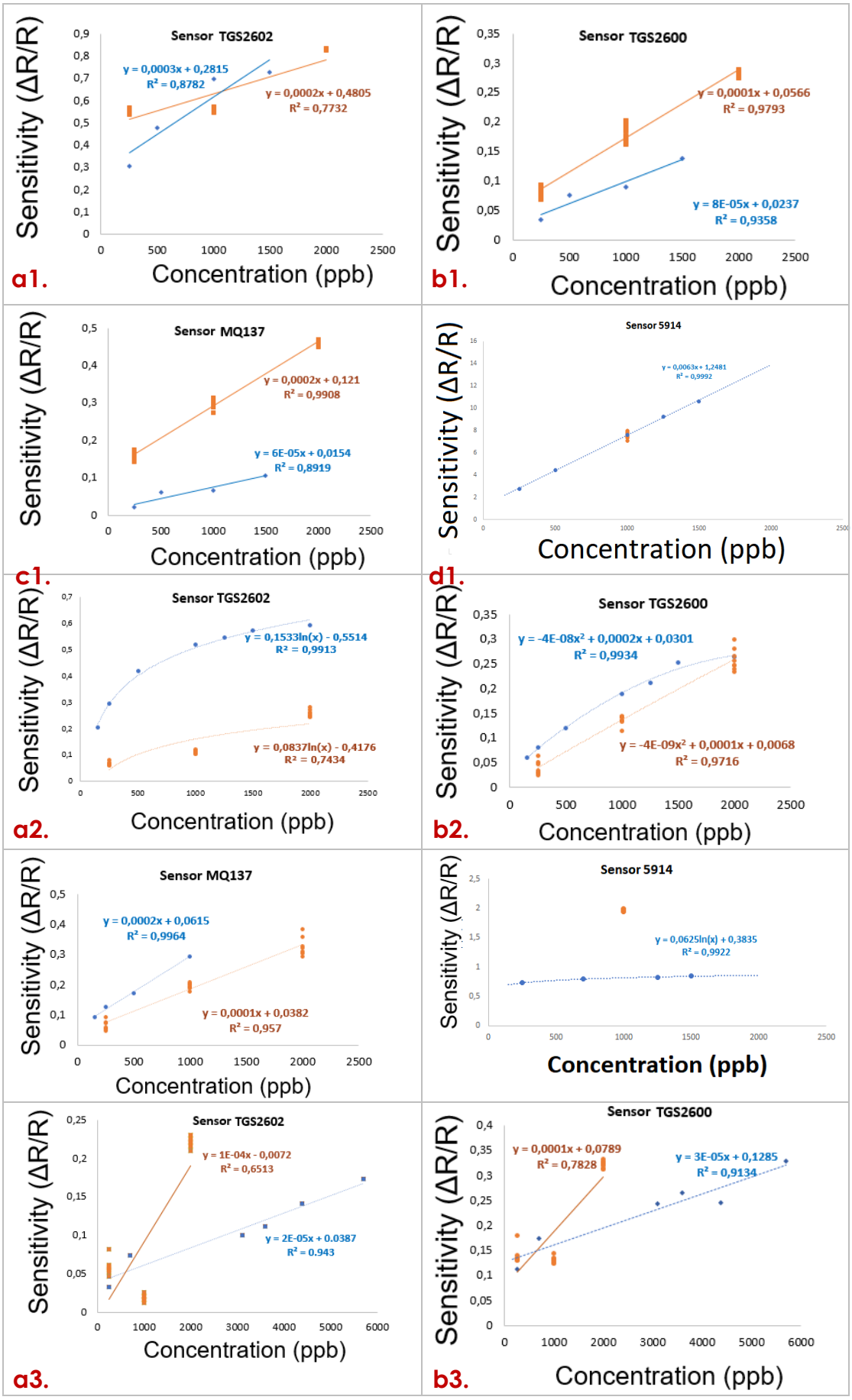
By looking at the data points of Figure 23, it can be understood that some kind of fitting can be implemented in order to estimate and ultimately predict the values of sensitivity for concentrations that have not been measured experimentally in the lab. In Figure 24 one can appreciate the fitting of the data from the "Sensitivity series" merged with the data collected from the "Reproducibility series". For what concerns the "Reproducibility series" the data correspond to multiple points of sensitivities that have been extracted for concentrations of 250ppb, 1000ppb and 2000ppb (for each gas) for board I and for concentrations of 1000ppb (for each gas) for board II. The Figures ai (i=1,2,3) correspond to sensor TGS2602, the Figures bi to TGS2600, the Figures ci. to MQ137, and the Figures di to 5914, where i=1 corresponds to acetone, i=2 corresponds to toluene and i=3 corresponds to benzene. On each Figure the equation of the each trendline is shown. The blue curves correspond to the fitted data from the

"Sensitivity Series" (RC plot/data, i.e Range of Concentration), while the orange curves correspond to the fitted data from the "Reproducibility Series" (R plot/data, i.e Reproduciblity).

For the case of acetone, and specifically for sensor TGS2602, even though both RC and R plots follow a linear trendline their respective intercept is pretty different, mainly attributed to the decrease in sensitivity of 1000pb concentration of acetone. It is worth noting that the R curves for this sensor show a lower sensitivity at 1000ppb of acetone than at 250 ppb. Since there has been a long time in between the R data of 1000ppb and the R data of 250ppb, those differences may be attributed to system effects (humidity, pressure fluctuations, drift etc). For sensors TGS2600 and MQ137 the slope of the trendline is almost the same for both RC and R plots. The difference in the intercept may be attributed to the drift effect. This means that the sensor interacts with acetone somewhat reliably and certainly suggests that a calibration is needed during the operation of the sensors to counteract the drift. For sensor 5419 there are R data only for 1000ppb. However, it is worth mentioning, that the reproducibility for the case of acetone is excellent even after an extended period of time.

For toluene, sensor TGS2602 seems to have a logarithmic trendline when it comes to RC plots. In the case of the R data the datafit has been realized in a logarithmic trendline of $R^2 = 0.7434$ (R^2: coefficient of determination, is a measure of the goodness of the fit) in order to match the trendline of the RC data. However, when the linear fitting was applied for the same R data, a higher R^2 of 0.9405 was given, which means that the linear fit is more appropriate. The change of the trendline might be due to the smaller number of R data compared to the RC data though. Sensor MQ137 presents a similar linear regressive characteristic for both RC and R plots. The slight difference in the slope is possibly introduced by the expansion of range of the RC data, while the difference in the intercept may have been introduced by the drift effect as already mentioned. Sensor TGS2600 presents a R^2 of 0.9934 for the RC polynomial trendline. However, when a linear fit was attempted for the RC data the R^2 was equal to 0.9561, which is still high enough and it is assumed that since the vast majority of the previous trendlines gives a linear trend, a linear fit is probably the right trendline to be chosen. Sensor 5914 presents a linear fit when it comes to the RC data but the R data do not much the fit, as they did in the case of acetone. Since the baseline drift did not affect the reproducibility of acetone it would be contradictory to assume that it affected the reproducibility of toluene, at least for this specific sensor. It can be deducted though that sensor 5914 is much more accurate and suitable for sensing acetone rather than toluene.

Similarly to toluene, the sensor 5914 seems to be also poorly suitable for the detection of benzene for the same reasons. In addition, the sensors TGS2602, TGSMQ137, and MQ137 show the exact behavior when comparing their respective RC and R trendlines. Specifically, one can observe a gradient shift, and an outlying behavior especially of the 1000pb R data.


Sensor TGS2602
y = 0,0003x + 0,2815
R² = 0,8782
y = 0,0002x + 0,4805
R² = 0,7732
Sensitivity (ΔR/R)
Concentration (ppb)
a1.
Sensor TGS2600
y = 0,0001x + 0,0566
R² = 0,9793
y = 8E-05x + 0,0237
R² = 0,9358
Sensitivity (ΔR/R)
Concentration (ppb)
b1.
Sensor MQ137
y = 0,0002x + 0,121
R² = 0,9908
y = 6E-05x + 0,0154
R² = 0,8919
Sensitivity (ΔR/R)
Concentration (ppb)
c1.
Sensor 5914
y = 0,0063x + 1,2481
R² = 0,9992
Sensitivity (ΔR/R)
Concentration (ppb)
d1.
Sensor TGS2602
y = 0,1533ln(x) - 0,5514
R² = 0,9913
y = 0,0837ln(x) - 0,4176
R² = 0,7434
Sensitivity (ΔR/R)
Concentration (ppb)
a2.
Sensor TGS2600
y = -4E-08x² + 0,0002x + 0,0301
R² = 0,9934
y = -4E-09x² + 0,0001x + 0,0068
R² = 0,9716
Sensitivity (ΔR/R)
Concentration (ppb)
b2.
Sensor MQ137
y = 0,0002x + 0,0615
R² = 0,9964
y = 0,0001x + 0,0382
R² = 0,957
Sensitivity (ΔR/R)
Concentration (ppb)
Sensor 5914
y = 0,0625ln(x) + 0,3835
R² = 0,9922
Sensitivity (ΔR/R)
Concentration (ppb)
Sensor TGS2602
y = 1E-04x - 0,0072
R² = 0,6513
y = 2E-05x + 0,0387
R² = 0,943
Sensitivity (ΔR/R)
Concentration (ppb)
a3.
Sensor TGS2600
y = 0,0001x + 0,0789
R² = 0,7828
y = 3E-05x + 0,1285
R² = 0,9134
Sensitivity (ΔR/R)
Concentration (ppb)
b3.

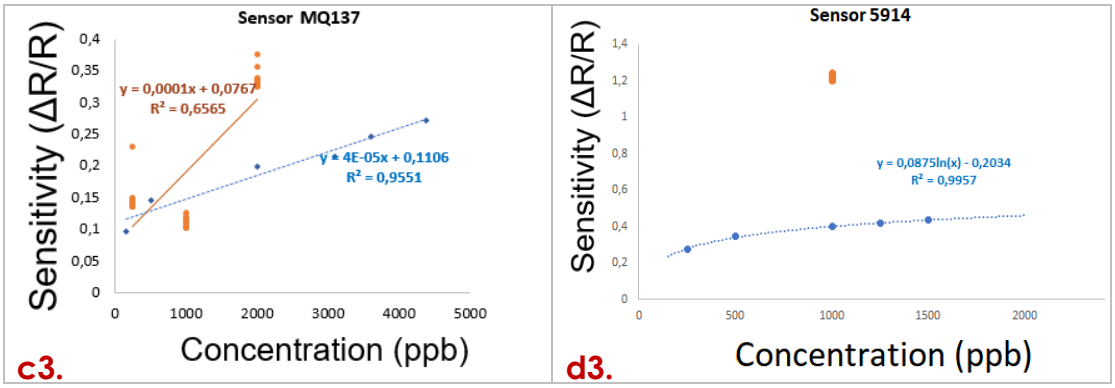


Figure 24. Fitting of Sensitivity vs concentration points and of their respective"reproducibility" points. Figures ai. correspond to TGS2602, Figures bi. to TGS2600, Figures ci. to MQ137, Figures di. to 5914, where for i=1 the results correspond to acetone, for i=2 the results correspond to toluene, and for i=3 the results correspond to benzene

3.1.4 Conclusions for commercial sensors

As it has been mentioned, all commercial sensors that exist on the market are not really labeled for the detection of specific gases. That means that the sensors used might be either unsuitable for acetone, toluene and benzene or those gases might even be poisonous for them. Based on the above results, it seems that sensor 5914 is a suitable sensor for what concerns the detection of acetone. The behavior of 5914 puts into question whether the effect of the drift is responsible for non-reproducible sensitivity values over an extensive period of time and rises the suspicions of whether the rest of the sensors are suitable enough for those specific gases. Nevertheless, by acquiring more R data in the future it might even be possible to obtain a trend of the respective shifts of the sensors as a function of time in order to impose corrections in the predictions. Under those conditions, as it has been shown, this specific set of sensors can operate in a complementary way, since TGS2602 and 5914 may give more accurate sensitivity response in the case of benzene over a wide range of concentrations in contrast to TGS2600 and MQ137 that fail to do so but can discriminate toluene and acetone in low concentrations in contrast to TGS2602 and 5914 that fail.

3.2 Cognitive Part

3.2.1 Automatic data analysis of commercial sensors

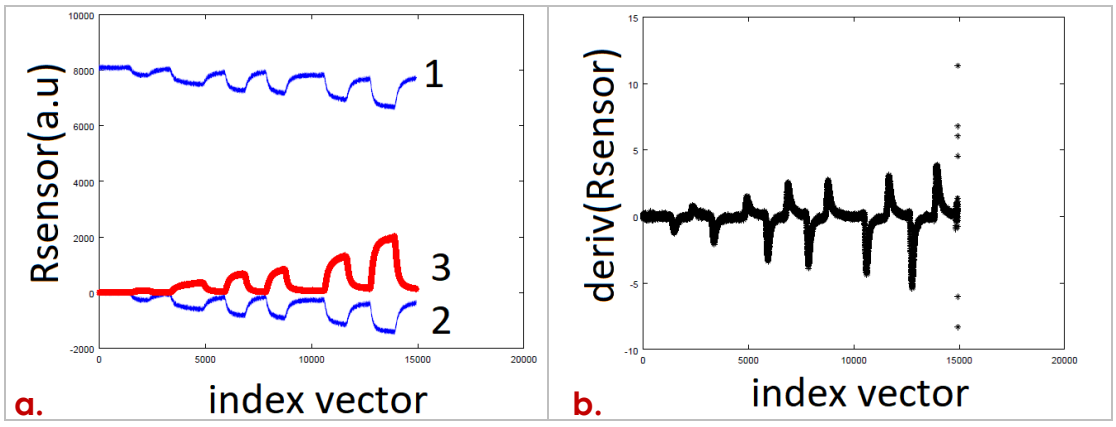
For the extraction of all the sensitivities, which have been obtained during the experiments that have been presented so far, a manual process was used. In other words, the author of the thesis would determine visually every peak in the graphs of Rsensor vs time after plotting the data with the help of a scientific graphing and data analysis software. In this case the software that has been used was Origin. After the visual detection of each peak, the Rgas and Rair values have been determined on the graph and their values were recorded and fed as input to the python script of Appendix A for the extraction of the sensitivities. Because of the long time

consumption of this process, an automatic determination of peaks and sensitivities was attempted. The script is written in Octave programming language and is included in Appendix B and C. Specifically the Appendix B includes the script that corresponds to the preprocess of the Rsensor data in order for them to be more "manageable" on the later steps of the process and Appendix C includes the actual determination of the peaks and the calculation of the sensitivity (process).

In Figure 25 the evolution of the Rsensor data is depicted during the preprocess of Appendix B. Plot number 1 of Figure 25a presents the data of Rsensor as they have been obtained from the experimental measurement, plotted on Octave. Even though the experimental setup provides voltage values that are afterwards converted into resistance values, the Rsensor values are going to be referred as raw data from now for sake of brevity. Plot number 2 in the same Figure shows the same raw data centered around zero and Plot number 3 presents the same data after they have undergone the so called "feature scaling', in order to smooth out the effect of excess noise and the evident baseline drift. In data processing, the last step is also known as data normalization and is generally performed during the data preprocessing step (2).

Specifically, the feature scaling consists in: 1. subtracting the mean value of the baseline from the raw data points 2. creating a new vector (i.e column of new points) with the squared values of the obtained data from the step before, 3. calculating the moving average of the new vector for every ten points and diving it by ten and by the variance of the baseline (variance is the expectation of the squared deviation of a random variable from its mean : $V(x) = E[(x-\mu)^2]$ (3)).

On a later step the derivative of every 200 points of plot number 3 was calculated and is shown in Figure 25b. The same data points have been lotted again in Figure 25c on top of the raw data that have been centered around zero. It becomes evident that the peaks of the derivative correspond to the raw data whose slope is the greatest and this means that the first peak corresponds to sensing, the second peak to recovery, the third to sensing and so on and so forth. In other words the peaks of an odd index correspond to sensing and the peaks of an even index correspond to recovery.



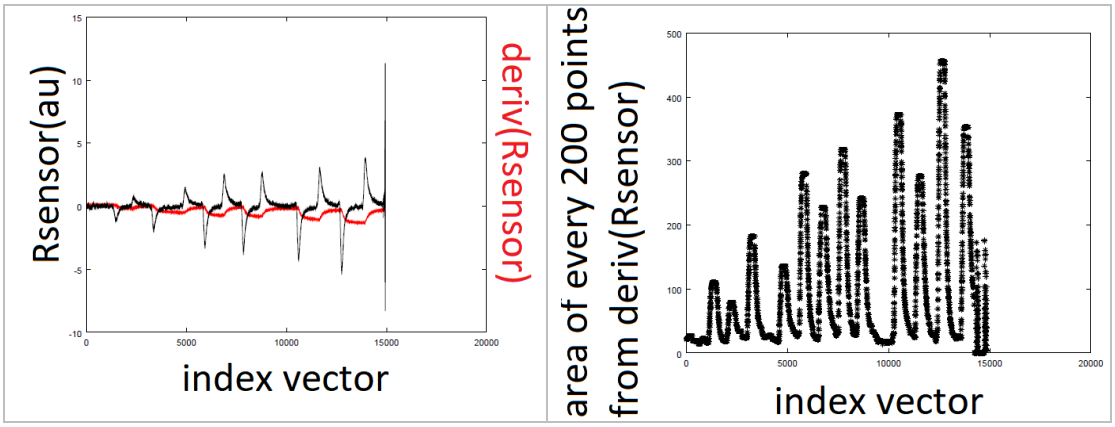


Figure 25. a. presents 1. the raw data , 2. The raw data centered around zero, 3. the raw data after feature scaling, b. presents the derivative of plot 3, c. presents the derivative of plot 3 on top of plot 2 , d. presents the calculation of area of every 200 points of the data of plot in figure b

Appendix C contains the script regarding the second step of the process, which includes the determination of the position of the peaks and the calculation of the sensitivity. For the determination of the peaks the following concept has been followed: A window of two hundred points capacity was sliding along all the data of the plot in Figure 25b and during each slide the respective area was calculated with a triangular approach. The results are plotted in Figure 25d and as it becomes obvious the maxima correspond to the positions of the peaks of the plot of Figure 25b. After identifying the position of the peaks, and keeping under consideration that odd index corresponds to sensing and an even index corresponds to recovery, the following concept was followed: a certain range of data before a peak of an odd index corresponds to values of the baseline (Rair) and a certain range of data before a peak of an even index correspond to values of the sensing (Rgas). By applying the equation 3 for every odd-even pair, the corresponding sensitivity can be calculated. In Table 3 the values of sensitivity are presented in the case of benzene for sensor TGS2602 as they have been calculated both manually and automatically. The discrepancy is due to the human error that was involved during the manual calculation but the discrepancy especially at the higher concentrations is also due to the short stabilized regions (baseline regions) that were obtained during the measurement that did not provide good Rgas values. However, the script seems to correspond fine and such problems can be eliminated by making the process of obtaining experimental data also automatic with long enough baseline regions.

Table 3

Conceentration (ppb)	250	700	3100	3600	4380	5700
Manual approach	0,033153	0.073971	0,09995	0,111948	0.141388	0.173431
Automatic approach	0.031945	0.070714	0.100003	0.109309	0.108264	0.134902

3.2.2 Machine learning for Temperature Controlled Operation (TCO) board

As it has already been mentioned the machine learning part is related to the experimental results obtained from the TCO mode. In Figure 26, one can appreciate some of the experimental data, as they have been delivered by the experimentalists in order for the author of the present thesis to work on the cognitive part of the project. It can be seen that the temperature (voltage ratio) follows a square waveform, where the voltage ratio drops from 100% to 30%, then again from 100% drops to 40%, then it drops to 50%, 60% and finally to 70%. The Rsensor values change when the temperature drops both in the case of just dry air and in the case of gas flow. In Figure 26 the gas corresponds to acetone of 250ppb. Each drop of temperature corresponds to an example. In the machine learning field the example "carries" the features that train the network. In our case, as it can be seen in Figure 26, there are five examples to train the network in order to recognize the case of acetone. In addition to those, four examples have been obtained for the case of benzene and five more examples for the case of toluene.

In other words, six different sets of experimental data have been delivered in total by the experimentalists: the variations of Rsensor due to the same pattern of voltage alternations presented in Figure 26, while 1. acetone of 250ppb is flowing, 2 and its respective dry air response, 3. while toluene of 250ppb is flowing, 4. and its respective dry air response, 5. while benzene of 250ppb is flowing and 6. Its respective dry air response. It's worth mentioning that there three different dry air responses and not just one, because every dry air measurement was obtained right before the gas measurement in order to obtain pair of measurements (1-2, 3-4, 5-6) whose drift is not that different from each other.

The goal of this part of the thesis is to find the best model that a neural network can provide in order to classify correctly unknown gas inputs. The neural network that was used for the purposes of this thesis, was scripted in python language and was not created by the author of the thesis. Nevertheless, the preprocess of the data that were fed to the network, the optimization of the model and the validation was carried out by the author. The PCA algorithm that has also been used, was also scripted and implemented by the author.

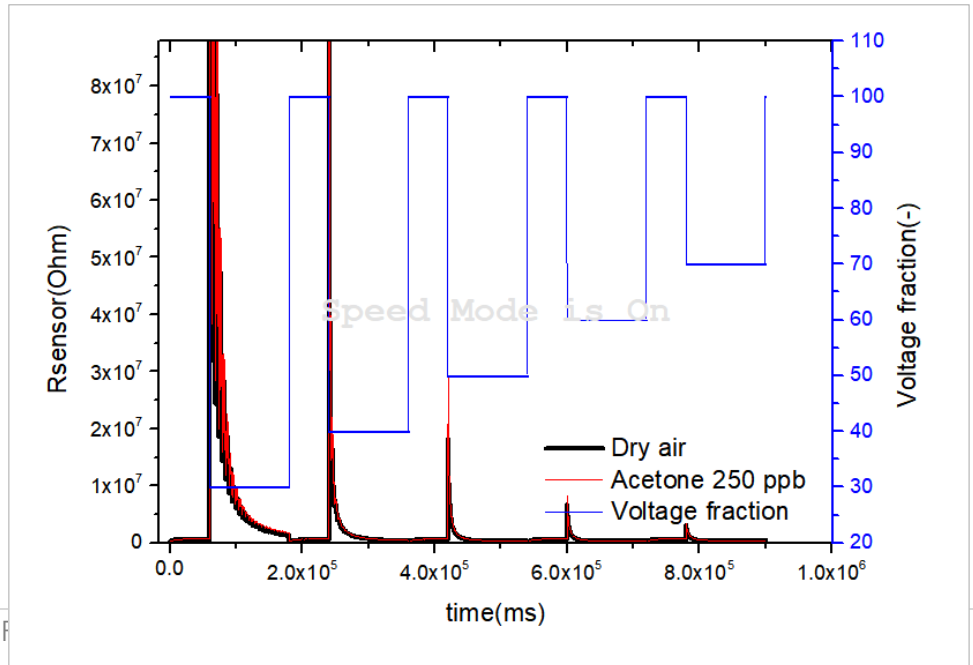


Figure 26. The variations of Rsensor for TGS8100 in a steady state operational mode (temperature follows a square waveform), while a constant flow of dry air is purging (black curve) and on a second step acetone of 250ppb (red curve)

In order to feed meaningful data to the neural network, a preprocess is always mandatory and is included in Appendix D. The concept of the preprocess that was followed was the following: 1. The point of every temperature drop was determined. 2. A certain number of Rgas and Rair data points were selected, that would correspond a bit before and a bit after each temperature drop. In that way the examples per gas were determined. 3. Each example was segmented in seven different regions (such segmentation can be appreciated in Figure 27 for the first example of acetone 250ppb). 4. The area of each segmented region was calculated and hence all areas per region of example constituted the features of each example. 5. All features were normalized, using the feature scaling technique already described, where the mean value of all features per example was used for the scaling. 6. All fourteen examples (five for acetone, five for benzene and four for toluene) and their corresponding features (seven features) were loaded in a common matrix with their respective labelling. This matrix is presented in Table 4.

The labeling serves the purpose of comparing the outputs of the neural network with the expected outputs. In Table 4 the expected outputs are presented per example. For instance, it can be noted that the column named "label Ace" equals 250 (in accordance to 250ppb) only for the very first five examples whose corresponding features are indeed the calculated areas of the seven regions in which all five data windows (data around the point of the temperature drop) have been divided.

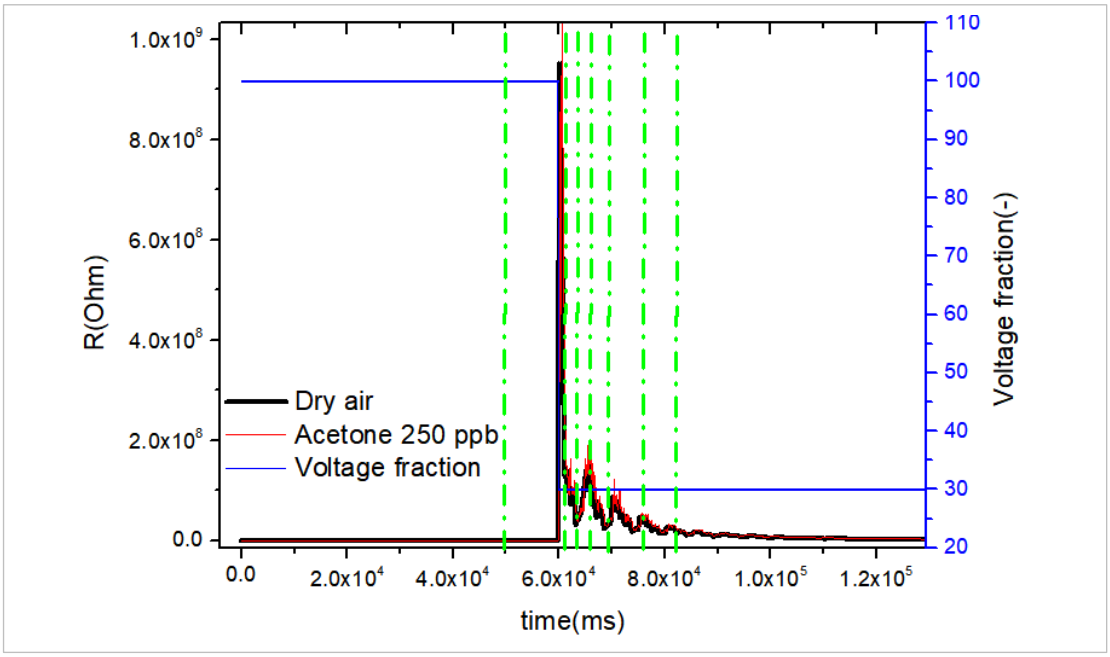


Figure 27. Segmentation of the first example of acetone 250ppb in seven regions, the area of which correspond to its respective features

Table 4

	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Label Ace	Label Ben	Label Tol
Example 1	0.0119	20.60	0.5086	12.217	11.134	11.006	0.8177	250	0	0
Example 2	0.0823	0.445	0.8986	0.7782	0.4912	0.4449	0.3528	250	0	0
Example 3	0.0897	16.65	0.2334	0.5532	0.4381	0.4101	0.3481	250	0	0
Example 4	0.0673	0.246	0.2402	0.4729	0.4577	0.3921	0.2459	250	0	0
Example 5	0.0696	0.324	0.5209	0.2639	0.3089	0.2250	0.0395	250	0	0
Example 6	12.433	0.313	15.954	0.7106	0.8340	0.8601	0.8240	0	250	0
Example 7	13.544	13.14	0.8677	13.478	13.593	13.298	13.521	0	250	0
Example 8	12.785	12.88	13.436	15.437	14.979	15.249	15.021	0	250	0
Example 9	0.9048	10.43	11.848	13.181	13.178	13.137	12.806	0	250	0
Example 10	0.6894	10.04	10.903	12.196	11.853	11.351	10.395	0	250	0
Example 11	11.576	0.129	13.515	10.887	10.928	11.232	12.731	0	0	250
Example 12	14.279	0.110	11.462	0.9019	10.768	10.462	11.921	0	0	250
Example 13	12.779	0.007	0.6750	0.4756	0.6257	0.6718	0.8331	0	0	250
Example 14	12.860	0.011	0.5071	0.3836	0.5894	0.7495	0.8959	0	0	250

The last step before feeding the data into the neural network corresponds to the shuffling of the rows of the matrix, which is depicted in Table 4, and its splitting into three different data sets that the neural network will use in order to extract the best model of classification. Specifically, 60% percent of data (examples with corresponding features and labels) was transferred in one file that was used for the training purposes of the network, 20% of the data was transferred in one file that was used for validation purposes, and the last 20% of the data was transferred in one file that was used for the actual testing of the model.

The two first data sets are used during the training of the neural network, where the first set is used to adjust the weights and the second set is used for minimizing overfitting problems, since the validation set plays the role of the internal "unknown inputs" of the neural network and by comparing the respective accuracies of the training and validation set the neural network will realize when to stop the training (3).

Several models have been tried out in order to conclude which model gives the most accurate classification, and the one that has been finally selected is the one of the following architecture characteristics : learning rate: 0.008, momentum: 0.9, Loss function: 0.0, Hidden Units: 11, and regularization parameter (Lambda): 0.3. The different models that have been tried considered variations of the above mentioned parameters.

In practice, the results can be appreciated by plotting the real labels (the ones shown in Table 4) and the ones derived from the neural network. A plot like that is depicted in Figure 28, where the blue points are the real labels/concentrations of the examples and the red points are the

classified labels/concentrations of the best model of the neural network. It can be observed that the network never failed to misclassify a gas for another gas but it did give some off values for the case of acetone and benzene. This error though can be justified, since the datasets that have been used are extremely small. However, the results suggest that this model is a good starting point and it provides plenty of room for improvement as soon as more data are fed into the network.

The chapter of the cognitive part closes with a PCA application. This suggests an extra step of implementation before feeding the data to the neural network. The normalized features provided a new set of features after a function (function sigma) was applied to them in order to reconstruct them in pairs of eigenvalues and eigenvectors, where the first represent the direction of variance and the second the variance of the examples. A script was written by the author of the thesis to determine the number of eigenvectors whose respective eigenvalues have variance greater than 99.5%, in order to maintain the important information and hence keep the most important features which can serve the purposes of the classification. It has been found that only three features were mandatory.

After the PCA was completed, the new three features were fed to the neural network for which the same initial values as the ones used for the case of the "best model" mentioned above, were used. In Figure 29 the plot of expected vs derived labels was plotted after PCA was performed. As expected, there is a dramatic increase of the error because of the limited number of initial data. In general, when the initial data are few, when PCA is applied it puts into risk eliminating features, which are actually beneficial for the classification.

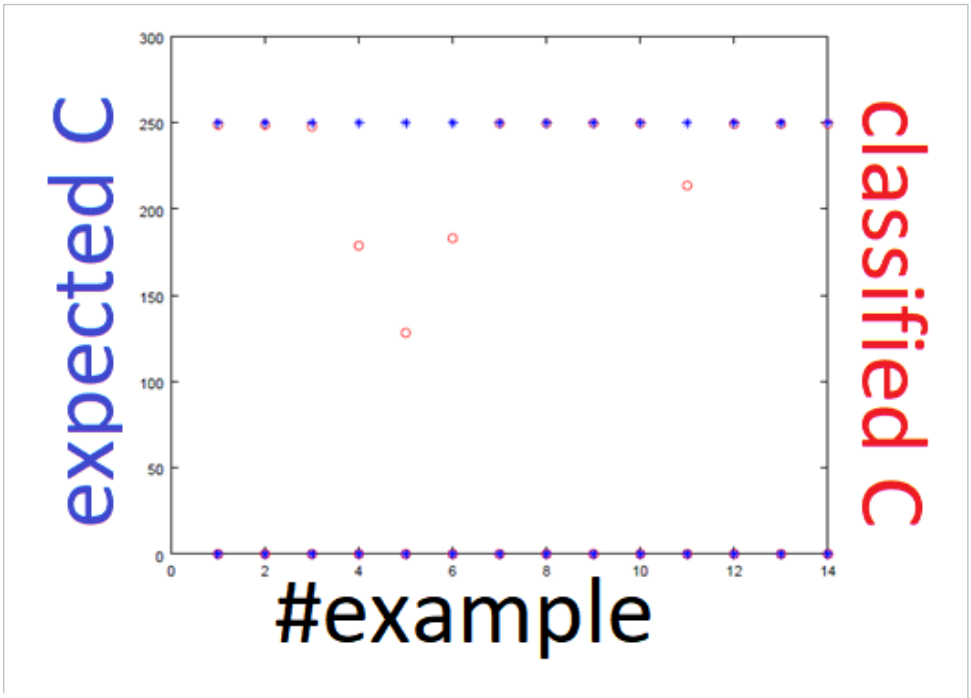


Figure 28. Expected vs derived labels from the best model obtained by the neural network with no PCA

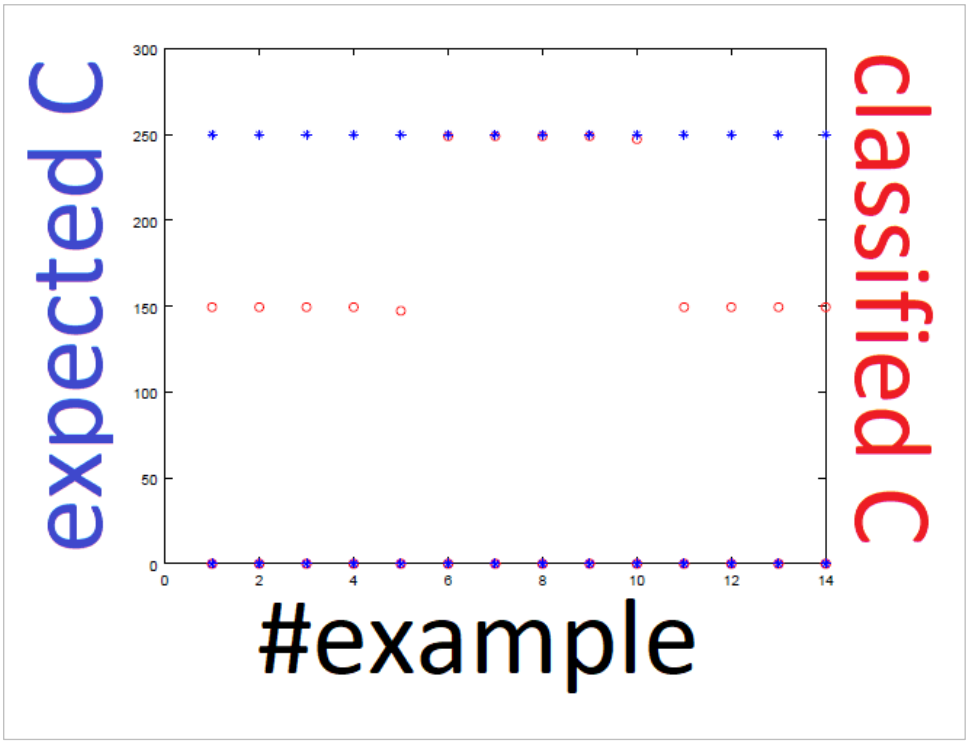


Figure 29. **Expected vs derived labels from the best model obtained by the neural network with PCA**

3.2.3 Conclusions for cognitive part

The cognitive part of this master thesis was divided into two parts. The first one was related to the automatic determination of the response of commercial sensors in "Single Temperature Operation" from the experimental data, and calculation of their respective sensitivities. The script was tested and it seemed to provide with pretty accurate results. The discrepancy of some results between the manual and the automatic approach was attributed to human error and mainly to the fact that that the high concentrations curves did not reach full stabilization of their baseline. The solution to this problem, as it has been mentioned, might be automatic overnight prolonged experiments to ensure the stabilization of the signal.

The second part was related to the machine learning implementation of the "Temperature Controlled Operation" mode. Even though the results were not satisfying enough, it has to be pointed out that this step constitutes a good starting point, since it was the first attempt of combining a cognitive part to the sensor project. From the cognitive part perspective a great number of data is mandatory to be obtained before establishing a robust classification model, while from the experimental perspective a series of further investigation is ought to be done in order to determine which patterns of temperature variation are the optimal in terms of high Rsensor response.

4.3 BIBLIOGRAPHY

(1). M. Leidinger, T. Sauerwald, W. Reimringer, G. Ventura, and A. Schütze. (2014)Selective detection of hazardous VOCs for indoor airquality applications using a virtual gas sensor array, JSSS, 3, 253–263

(2) Derived from https://en.wikipedia.org/wiki/Feature_scaling

(3) Derived from https://stackoverflow.com/questions/2976452

Appendix A

Python script for voltage-resistance conversions

```
print(' !!! IF the suply is not 1.8V, change the code! !!!! Better
use the average code before this one  !!! ')

# open the file you want to modify

file_name = input('Give the name of the txt file you want to modify
with no extention')

RL1,RL2,RL3,RL4,RL5,RL6 = input('Give the values of load resistances
in order').split()

RL =[RL1,RL2,RL3,RL4,RL5,RL6 ]

file_name_2= file_name + '.txt'

file_open = open(file_name_2, 'r')

time_list, V1_list, V2_list, V3_list, V4_list,V5_list,V6_list = ([]
for i in range(7))

A=[time_list, V1_list, V2_list, V3_list, V4_list,V5_list,V6_list]

for line in file_open:

 line_of_strings = line.split()

 for x in range(7):
```

```
  B=line_of_strings[x]

  A[x].append(B)

time_list, V1_list, V2_list, V3_list, V4_list,V5_list,V6_list=[A[i]
for i in range(7)]

# Convert the list from strings to floating numbers

  L=[V1_list,V2_list,V3_list,V4_list,V5_list,V6_list]

K=[]

for y in L:

 y = [float(i) for i in y]

 K.append(y)

 V1_list_float, V2_list_float, V3_list_float, V4_list_float,
V5_list_float, V6_list_float = [K[i] for i in range(6)]

# Transformation for every value of list

M = [V1_list_float, V2_list_float, V3_list_float, V4_list_float,
V5_list_float, V6_list_float]

N = []

for j in range(6):

  M[j] = ['%.8f' %((float(RL[j]))*(1.8/(i/1000)-1))  for i in M[j]]

  N.append(M[j])

 R1_list, R2_list,R3_list,R4_list,R5_list,R6_list= [N[x] for x in
range(6)]

# Make list ready for copying to a file

x=file_name + '_resistances.txt'

file = open(x, "w")

for value in range(len(time_list)):

 file.write(str(time_list[value]) + " " + str(R1_list[value]) + " " +
str(R2_list[value])+ " " + str(R3_list[value])+ " " +
str(R4_list[value])+ " " + str(R5_list[value])+ " " +
str(R6_list[value])+"\n")

file.close
```

Appendix B

Octave script for preprocess of Rsensor values in Single Temperature mode

```
function[maxima]=asimina2(thresh)
close all;
%clear all;
dat = dlmread('benzene_average_resistances.txt');

% initialize constants
```

```
num = 2; %from 2-7.
stPt = 1;
endPt = 1390;
amp = 0.07 ;
height = 8000 ; % the distance between raw-data's baseline and 0(0,0)
n_grad = 200 ; % the number of points for which the gradient is
calculated
step = 100 ; % the step with which I choose to reduce data of gy
add_step = 10; %the step in the window of each the peak detection is
taking place
height_range = 0.3 ; % its the precentage in the range of which two
neighboring peaks are con
peak_height_threshold = 200; % all maxima found should be higher than
a threshold or else the max corresponds to noise


%%Format like Ng ML course

%%%%% ---- Preprocessing ---- %%%%%

%Visualize data, first column is the time point and the others are
data
%points in elements of resistance. Sensitivity = [R1-R2]/R1


dat_holder = dat(:,num); %num changes by sensor insert accurate
sensor number
dat_baseline = dat_holder(stPt:endPt); %fill in with values according
to gas, visualization and Asimina.
dat_var = var(dat_baseline);
mean(dat_baseline);
dat_centered = dat_holder - mean(dat_baseline); %Basically you center
the graph arround zero

N = 10; %reduce_points size we've chosen.
dat_proc = (dat_centered.*dat_centered);
% moving average
gy=zeros(1,length(dat_proc)-N+1);
for i =1:N
gy = gy + dat_proc(i:end-N+i)';
end
gy=gy/(N*dat_var); %feature scaling
R = (gy>thresh);   % the figure of merit for concerns the smoothing
of the graph
datb = dat .- height;
grad_gy2=zeros(length(gy),1);
for i=n_grad:length(gy);
grad_gy2(i-n_grad+1)=(gy(i)-gy(i-n_grad+1)/n_grad)  ; % gradient of
processed between every n_grad points
end;
```

Appendix C

Octave script for determination of peaks and sensitivity

```
clear all;
close all;

load('autom.mat');
n = 200; %window size
```

```
%datbpr is the derivative of the original response curve of VOC's
exposure
%to sensor.
rm = 0;
limit=500; %ending point for calculation of baseline average;
bias= 100; %number of points to adjust the real pea position to the
predicted one
diff =30;  % diff = point(i+1) - point(i), where both points are
points of the baseline if the diff is small enough

datbpr =datbpr(1:end-500);%take out the noise by cutting off last
hundred pts
% stem(datrp)

figure;
plot(datbpr, 'k*')


bias = 100;
datbpr=datbpr(1:length(datbpr)-rm);

for i=1:length(datbpr)-n ;
     datfnd=abs(datbpr(i:i+n-1)); %flooring it is no good
     datr(i)=abs(0.5*n*max(datfnd));
end

figure;
plot(datr,'k*');

datrp =datr;

base=datrp(1);

mrk = 0;
j = 1;

%this code works well in detecting peaks of the values for triangle
method

for i = 2:length(datrp)
    hld = datrp(i);
    if(hld == datrp(i-1))
        mrk = mrk+1;
        if(mrk == n-1)&& (hld > base*2) %multiplying the baselne by 2
improves threshold
            val(j) = i; %i-n+1
            j = j+1;
        end
    else mrk = 0;
    end
end

numpeak= length(val)/2;


%&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&


% determing the value of baseline

baseline_sum=0;
```

```
for i =1:1:length(val)/2   % val(i) is the index for the ith peak in
%the datb !!!!! I check the baseline before peak=1,3,5,...11
window = datb(val(2*i-1)-limit-bias:1:val(2*i-1)- bias);   % I choose
a window of #limit# points before   peak=1,3,5,...11
a=1; % variable that breaks the for loop in case elseif is valid


  for j = 1:length(window) ;
  i
  j
  hld = window(j)
  hld1=window(1)
  %window(j+1)
  %x= hld - hld1
   x= hld -  hld1
   abs(x)
  if (abs(x) < diff) && a==1
  baseline_sum = baseline_sum + window(j)
   if (j==length(window)-1);  % calculate the average of baseline
ONLY after the sum is complete for all j
   baseline_average(i) = baseline_sum/(length(window))
   baseline_sum=0;
   endif
  else
  baseline_average(i)= baseline_average(i-1)
  baseline_sum=0;
  a=0;
  endif
  end;
sens(i)= ((baseline_average(i) - datb(val(i*2)))/baseline_average(i))
 end
```

Appendix D

Octave script for preprocess of Rsensor values in Steady State Operational Mode

```
clear all;

close all;

tempdat =dlmread('acetone_250_ppb_square_waveform_100%.txt');

time_samp = 1000;

step_samp = 10;

time_air = tempdat(:,1);   % time

time_ace = tempdat(:,4);   % NO NEED TO BE USED % time is the same

rest_air = tempdat(:,2);   % R of air

rest_ace = tempdat(:,5);   % R of acetone

volt_air = tempdat(:,3);   % V% of air

volt_ace = tempdat(:,6);   % NO NEED TO BE USED % V% is the same

total_points =400;
```

```
points_bf_peak = total_points*0.25          % The number of points
taken after the Temperature drop

points_aft_peak =total_points*0.75;       % The number of points
taken before the Temperature drop

%tair_sec = time_air/time_samp;

%tace_sec = time_ace/time_samp;

figure

plot(time_air, rest_air, 'r*');

hold on

plot(time_air, rest_ace, 'k*');

rest_ratio = rest_air./rest_ace;   % Rair/Rgas equivance of Ggas/Gair
on the paper


figure

plot(time_air, rest_ratio, 'b*');

figure

plot(time_air, volt_ace, 'g');



rec = 1;

for i =1: length(volt_ace)-1;

 volt_ace(i+1);

 volt_ace(i);

 volt_ace(i+1)-volt_ace(i);

 if (volt_ace(i+1)-volt_ace(i) <=0);

 split_volt_examples(i,rec) = (volt_ace(i));

 split_rest_air_examples(i,rec)= (rest_air(i));

 split_rest_ace_examples(i,rec)= (rest_ace(i));

 split_rest_ratio_examples(i,rec)= (rest_ratio(i));

 else

 rec = rec +1;

 end

 end

figure

plot(split_volt_examples(:,1), 'g');

hold on

plot(split_rest_ace_examples(:,1).*1e-6, 'k');
```

```
figure

plot(split_volt_examples(:,1), 'g');

hold on

plot(split_rest_ratio_examples(:,1), 'r*');

disp('The number of examples correspond to Resistance jumps  because
Temperature decreased from Tmax till Tmin. The number such examples
is:')

disp(columns(split_volt_examples)-1)

for j =1:columns(split_volt_examples)-1;

 for i =1: length(split_volt_examples)-1;

   if  (split_volt_examples(i+1,j)-split_volt_examples(i,j) <0) ;

       if  (split_volt_examples(i+1,j) != 0);

 window_volt_per_example(:,j) =  split_volt_examples(i-
points_bf_peak:i+points_aft_peak,j);

 window_rest_air_per_example(:,j) = split_rest_air_examples(i-
points_bf_peak:i+points_aft_peak,j);

 window_rest_ace_per_example(:,j) = split_rest_ace_examples(i-
points_bf_peak:i+points_aft_peak,j);

 window_rest_ratio_per_example(:,j) = split_rest_ratio_examples(i-
points_bf_peak:i+points_aft_peak,j);

 end

 end

  end

 end

figure

plot(window_volt_per_example(:,1), 'g');

hold on

plot(window_rest_air_per_example(:,1).*1e-6, 'k*');

points_bf_peak

length_after_peak = (length(window_volt_per_example)- points_bf_peak)

percentage1       =  round(length_after_peak*0.083)

sec_div           =  points_bf_peak + percentage1
%the first division includes the 25%  of the points after the Temp
drop

third_div         =  sec_div   + percentage1
%the second division includes another 25%  of the points after the
first division

percentage2       =  round(length_after_peak*0.167)

fort_div          =  third_div + percentage2

fif_div           =  fort_div + percentage2
```

```
six_div           =  fif_div   + percentage2


for i=1:columns(split_volt_examples)-1;
x=window_volt_per_example(:,i);
y=window_rest_air_per_example(:,i);
z=window_rest_ace_per_example(:,i);
w=window_rest_ratio_per_example(:,i);
  %first division                    %only the points before the peak
(# of point is 100)
divx1(:,i)=x(1:points_bf_peak);
divy1(:,i)=y(1:points_bf_peak);
divz1(:,i)=z(1:points_bf_peak);
divw1(:,i)=w(1:points_bf_peak);
  %second division                   (# of point is 25)
 divx2(:,i)=x(points_bf_peak+1:sec_div );
divy2(:,i)=y(points_bf_peak+1:sec_div );
divz2(:,i)=z(points_bf_peak+1:sec_div );
divw2(:,i)=w(points_bf_peak+1:sec_div );
   %third division                     (# of point is 25)
divx3(:,i)=x(sec_div+1:third_div );
divy3(:,i)=y(sec_div+1:third_div );
divz3(:,i)=z(sec_div+1:third_div );
divw3(:,i)=w(sec_div+1:third_div );
    %fourth division                    (# of point is 50)
divx4(:,i)=x(third_div+1 : fort_div  );
divy4(:,i)=y(third_div+1 : fort_div  );
divz4(:,i)=z(third_div+1 : fort_div  );
divw4(:,i)=w(third_div+1 : fort_div  );
     %fifth division                       (# of point is 50)
divx5(:,i)=x(fort_div+1:fif_div );
divy5(:,i)=y(fort_div+1:fif_div );
divz5(:,i)=z(fort_div+1:fif_div );
divw5(:,i)=w(fort_div+1:fif_div );
    %sixth division                       (# of point is 50)
divx6(:,i)=x(fif_div + 1:six_div);
divy6(:,i)=y(fif_div + 1:six_div);
```

```
divz6(:,i)=z(fif_div + 1:six_div);
divw6(:,i)=w(fif_div + 1:six_div);
  %seventh division                (# of point is all the rest)
divx7(:,i)=x(six_div+1:end);
divy7(:,i)=y(six_div+1:end);
divz7(:,i)=z(six_div+1:end);
divw7(:,i)=w(six_div+1:end);
end
a=length(divx1)
b=length(divx2)
c=length(divx3)
d=length(divx4)
e=length(divx5)
f=length(divx6)
g=length(divx7)
deltax = [a,b,c,d,e,f,g]    %contains the lenghts of all features
j =0;
for i = 1: length(deltax)
  start = j+1;
  stop = j+deltax(i);
%  sum(window_rest_ratio_per_example((j+1):(j+deltax(i),:))
feat(i,:)
=sum(window_rest_ratio_per_example(start:stop,:))*deltax(i);
j = stop;
end
feat_acetone= feat'; %preserve the way we represent training examples
num_rows=size((feat_acetone),1)
lab_acetone= zeros(3,num_rows)'
lab_acetone(:,1)=250
```