Bayesian Inference using Spike Latency Codes for Quantification of Health Endangering Formaldehyde

Muhammad Hassan and Amine Bermak

School of Engineering
Hong Kong University of Science & Technology, Hong Kong
Emails: mhassan@ust.hk, eebermak@ece.ust.hk

Amine Ait Si Ali and Abbes Amira

College of Engineering
Qatar University, Doha, Qatar
Emails: amine.aitsiali@qu.edu.qa, abbes.amira@qu.edu.qa

***Abstract*—Recently, the exposure to formaldehyde has appeared as a major concern since it is listed as a human carcinogen. Conventional methods for its long-term monitoring are not feasible due to their high operational cost, long analysis time and the requirement of specialized equipment and staff. In this paper, we use an electronic nose, containing an array of commercially available Figaro gas sensors, to estimate formaldehyde concentration. A hardware friendly bio-inspired spike latency coding scheme has recently been employed for gas classification by using relative time between spikes. We use this scheme to estimate formaldehyde concentration by utilizing absolute spike timings. However, there is no straightforward relationship between the spike latency and the formaldehyde concentration. Instead, stochastic variability in the sensor array response, corresponding to repeated exposure to the same formaldehyde concentration, implies that latency patterns of the sensor array encode probability distribution over the formaldehyde strength. We use a Bayesian inference approach to estimate the formaldehyde concentration, and its performance is successfully validated by acquiring data for formaldehyde with our sensor array at twenty different concentrations in the laboratory environment.**

Keywords–Formaldehyde exposure; Sensor array; Spike latency coding; Bayesian inference.

I. Introduction

Formaldehyde (CH_2O) is one of the most ubiquitous and reactive aldehydes in the environment. It is a colourless and rapidly polymerizing gas at room temperature and is widely used in consumer products to protect them from spoilage by microbial contamination. It can also be found in pressed wood products, tobacco smoke and fuel burning appliances [1]. Recently, formaldehyde received great attention when it was considered as a human carcinogen in the report of International Agency for Research on Cancer [2]. This was based on sufficient evidence of carcinogenicity from studies of human cancers and exposure to formaldehyde. Higher concentration levels of formaldehyde in the indoor environment pose a serious health hazard to occupants of buildings. A recent study [3] reported increased concentration levels in urban areas. This alarming situation highlights the importance of formaldehyde monitoring with a low cost and robust solution on a long-term basis for healthy living. Unfortunately, traditional methods [3]-[6] can not be used for the long-term monitoring of formaldehyde because specialized equipment and staff are required for the analysis of air samples collected from the area being monitored. Moreover, the cost and analysis time associated with these methods is very high.

An electronic nose system, containing an array of gas sensors, emerged as a successful platform for the fast identification of gases in the last two decades and it is targeted for many applications like food quality checking [7], diseases diagnosis [8], bacteria identification [9], environmental monitoring [10], beverages classification [11], paper quality inspection [12] and identification of health endangering indoor gases [15]. In this paper, we use an electronic nose system, containing an array of six commercially available Figaro gas sensors, to acquire the signature of CH_2O at twenty different concentration values spanning from 0 to 5 parts per million (ppm).

Motivated by the recent experimental findings in the field of neuroscience which report a logarithmic relationship between the odor concentration and the spike latency of the mitral cell [13], a logarithmic time domain scheme is used for gas classification by translating a sensor array response into a spike latency pattern [14]. Hardware friendly rank order based classifiers are developed for gas identification by using this technique [14]-[17]. In these classifiers, the relative time between spikes is utilized to distinguish gases. We use this scheme to retrieve concentration information by using absolute spike latencies. However, there is no straightforward relationship between the absolute spike latency and the formaldehyde concentration as there is in the rank order based classifiers where the change in relative times between spikes does not change the classification performance as long as their temporal order is not changed. Generally, gas sensors exhibit randomness in their responses due to multiple reasons and as a result, stochastic variability is observed in the latency patterns.

In this paper, we use a probabilistic inference approach [18] to deal with this randomness in the latency patterns in order to estimate CH_2O concentration. Probabilistic models have been successfully used in neuroscience to build computational theories for perception and action [19]. There are two major steps in this approach. The first step is to learn the probability encoding model or the tuning curve for the spike latency patterns at each concentration value of the CH_2O from the experimental data obtained through the sensor array. The second step is to use a bayesian decoding model to estimate the formaldehyde concentration for a new test latency pattern by using the probability encoding model. The performance of this approach is evaluated by acquiring the data of twenty different concentration values of CH_2O in the laboratory environment.

The paper is organized as follows. Section II explains probabilistic inference approach for CH_2O concentration estimation. Section III describes the experimental setup for data

 ISBN: 978-1-61208-426-8

acquisition and evaluates the performance of our approach. Finally, the conclusion is drafted in Section IV.

II. PROBABILISTIC INFERENCE

A logarithmic time-domain encoding scheme has been used for gas identification in rank order based classifiers [14]-[17]. In these classifiers, the spike latency l_i of the *i*-th sensor corresponding to a target gas is represented as

$$l_i = \frac{log\ x_i}{m_i} \tag{1}$$

where x_i denotes the sensitivity of the the sensor *i* and m_i is a sensor dependent parameter which is extracted through linear regression between the average log sensitivity of the sensors across the array as an explanatory variable and the sensitivity of the *i*-th sensor as an output variable. This spike latency carries information about the gas identity and its concentration. In rank order based classifiers, a temporal sequence of spikes referred to as a rank order is used for gas identification. Absolute spike latency is shifted with the change in the concentration but the temporal sequence of spikes remains fixed. In this paper, we use this logarithmic time-domain encoding scheme to retrieve concentration information by using absolute spike latency.

The potential challenge with this scheme is that the gas sensors usually exhibit randomness in their responses because of drift and as a result, stochastic variability is observed in the latency patterns. In this paper, we present a probabilistic inference approach to retrieve concentration information from the random latency patterns. Probabilistic approaches have been successfully used in developing computational paradigms for biological sensory systems [19].

The main objective of using probabilistic inference is to find the most probable concentration value of the new test latency pattern by learning the distribution of latency patterns corresponding to each concentration value from available measurements taken with the electronic nose. Let us consider the following notations for this probabilistic inference problem: suppose we have a set of concentrations $\boldsymbol{c}$=$\{c_j\}$ and latency patterns obtained through the experiments where each latency pattern is denoted as $\boldsymbol{l}$=$\{l_1, l_2, ..., l_n\}$, where l_i represents the latency of the *i*-th sensor.

Probabilistic inference is a two step process [19]. In the first step, we learn a model fitting that captures the mapping from $\boldsymbol{l}$ to $\boldsymbol{c}$ from available sensor array measurements. In the second step, we use bayesian decoding to estimate the concentration value c_i from the new observed latency pattern $\boldsymbol{l}$.

In order to learn model fitting, we need to know the distribution or probability encoding model of the latency pattern conditioned on the CH_2O odor intensity. With a particular model, parameterized by a vector θ, we can use maximum likelihood (ML) to obtain the optimal estimate $\hat{\theta}$ for which the latency patterns are most likely

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\boldsymbol{l}|\boldsymbol{c}, \theta) \tag{2}$$

In our application, we assume that latency patterns follow a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and hence second order statistics, that is, mean and covariance, is sufficient to learn this distribution. We use ML to estimate these parameters from sensor array measurements. If $\boldsymbol{m}$ is the ML estimate of the true mean and $\boldsymbol{S}$ is the ML estimate of the true covariance matrix then the conditional density of latency patterns with a given CH_2O odor intensity (c_i) is given by

$$p(\boldsymbol{l}|c_i) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{S}_i|^{1/2}}\ exp[-\frac{1}{2}(\boldsymbol{l}-\boldsymbol{m}_i)^T\boldsymbol{S}_i(\boldsymbol{l}-\boldsymbol{m}_i)] \tag{3}$$

In order to estimate a new latency pattern, Bayesian decoding is used to compute the posterior probability p($c_i|\boldsymbol{l}$) of every concentration value c_i in the set with a given latency pattern $\boldsymbol{l}$. It can be described as

$$p(c_i|\boldsymbol{l}) = \frac{p(\boldsymbol{l}|c_i)p(c_i)}{p(\boldsymbol{l})} \tag{4}$$

Finally, the concentration with maximum posterior probability is selected as an estimated concentration of the new observed latency pattern.

$$\hat{c} = \underset{i}{\operatorname{argmax}}\ p(c_i|\boldsymbol{l}) \tag{5}$$

III. EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

We use six commercially available Figaro gas sensors to build an array for CH_2O concentration estimation. The description of these sensors is listed in Table I. The experimental setup for acquiring the response of CH_2O at different concentrations is shown in Figure 1. The sensor array is embedded in a glass container with an inlet valve for CH_2O exposure and outlet valve for its outflow. The cylinders of CH_2O and dry air are connected to mass flow controllers (MFCs) which are used to control the CH_2O concentration by mixing it with air in different proportions. A computer with a data acquisition board is used for MFCs programming to achieve the desired concentration of CH_2O and for acquiring the response of the sensor array.

TABLE I. GAS SENSORS USED TO ESTIMATE CH_2O CONCENTRATION.

Sensor	*Target Compounds*
TGS 826	Ammonia
TGS 2600	Air contaminants
TGS 2602	Volatile organic compounds
TGS 2610	Liquefied petroleum gas
TGS 2611	Methane
TGS 2620	Solvent vapors

We expose the sensor array to twenty different concentration values of CH_2O in the range between 0.25 ppm to 5 ppm with a 0.25 ppm increment step. The sensor array is exposed to CH_2O for 500 seconds to obtain its response and then dry air is used for 750 seconds to recover the baseline response, i.e., the response without CH_2O vapors. All the sensors in the array respond to the target concentrations of CH_2O with different values of sensitivity. A typical response of a sensor in the array to CH_2O at ten different concentrations is shown in Figure 2. In the figure, from left to right the concentration is increased which results in decreased sensor resistance. Notice that drift also appears, i.e., the sensor does not recover to its original state (baseline response) during dry air exposure.

Drift is the major issue with current gas sensor technology, which may occur due to changes in operational conditions,

 ISBN: 978-1-61208-426-8

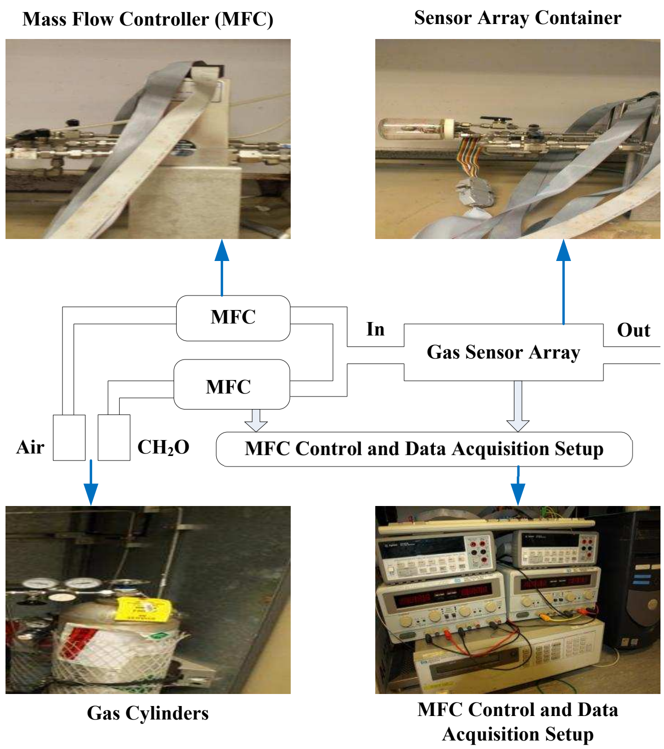

Figure 1. An Experimental setup to acquire the sensor array response to CH_2O exposure at different concentrations.

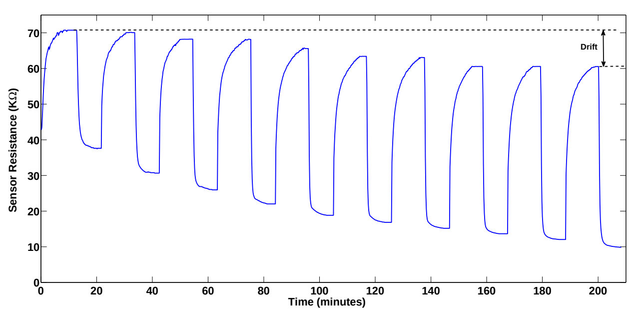

Figure 2. Typical sensor response to formaldehyde at ten different concentrations.

poisoning, and aging. Stochastic variability is observed as a result of this drift. From these resistance values, the sensitivity of each sensor is computed by dividing the steady state sensor resistance during the gas exposure by the baseline resistance. Regression coefficients m_i of each sensor are computed through linear regression, as shown in Figure 3. These parameters are used to transform the sensitivity pattern of the sensor array into a spike latency code (SLC) or pattern.

The Bayesian inference approach is applied on these spike latency codes in order to estimate the CH_2O concentration, and a 92.75$\%$ performance is achieved. This is slightly higher as compared to other state of the art methods, which include Gaussian mixture models (GMM), multi-layer perceptron (MLP) and support vector machines (SVM) with linear and radial basis function (RBF) kernel. The performances of all these methods are summarized in Table II. The Bayesian inference approach with spike latency codes not only performs

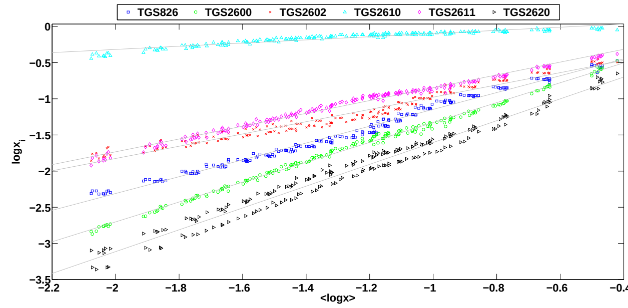

Figure 3. Extraction of regression coefficients for each sensor through linear regression.

better but it also does not require any tuning of parameters, which is extensively used in other methods.

TABLE II. PERFORMANCE COMPARISON OF CH_2O QUANTIFICATION ALGORITHMS.

Classification method	*Classification Performance (%)*
GMM	91.25
MLP	89.25
SVM (Lin)	88.5
SVM (RBF)	92.5
Bayesian Inference with SLC	92.75

IV. CONCLUSION

In this paper, we developed a low cost and compact solution to estimate the concentration of health endangering formaldehyde in the indoor environment by integrating commercially available gas sensors. Spike latency codes are used in the hardware friendly rank order based classifiers for gas classification. We use these codes to retrieve concentration information. Stochastic variability is observed in the spike latency codes due to inherent issues in the existing gas sensor technology. A probabilistic inference approach is used with spike latency codes to reliably estimate formaldehyde concentration. This scheme requires no manual tuning of the parameters as compared to other commonly used state of the art methods.

ACKNOWLEDGMENT

This paper was made possible by National Priorities Research Program (NPRP) grant No. 5 - 080 - 2 - 028 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] M. N. Indang , A. S. Abdulamir , A. A. Bakar , A. B. Salleh, Y. H. Lee, Y. N. Azah, “A review: methods of determination of health-endangering formaldehyde in diet,” Res. J. Pharmacology, vol. 3, no. 2, 2009, pp. 31-47.

[2] International Agency for Research on Cancer, “IARC classifies formaldehyde as carcinogenic to humans,” Press release. Jun. 2004, URL: http://www.iarc.fr/en/media-centre/pr/2004/pr153.html [accessed: 2015-07-07].

[3] P. Guo, K. Yokoyama, F. Piao, K. Sakai, “Sick building syndrome by indoor air pollution in Dalian,” Int. J. Environ. Res. Public Health, no. 2, Apr. 2013, pp. 1489-1504.

 ISBN: 978-1-61208-426-8

[4] T. Salthammer, S. Mentese, R. Marutzky, “Formaldehyde in the indoor environment,” Chemical Reviews, vol. 110, no. 4, Apr. 2010, pp. 2536-2572,

[5] J. G. Dojahn, W. E. Wentworth, and S. D. Stearns, “Characterization of formaldehyde by gas chromatography using multiple pulsed-discharge photoionization detectors and a flame ionization detector,” J. Chromatogr. Science, vol. 39, no. 2, Feb. 2001, pp. 54-58.

[6] H. L. Pinheiro, M. V. de Andrade, P. A. de Paula Pereira, J. B. de Andrade, “Spectrofluorimetric determination of formaldehyde in air after collection onto silica cartridges coated with Fluoral P,” Microchem. Journal, vol. 78, no. 1, Sep. 2004, pp. 15-20.

[7] S. Capone, P. Siciliano, F. Quaranta, R. Rella, M. Epifani, L. Vasanelli, “Analysis of vapours and foods by means of an electronic nose based on a sol-gel metal oxide sensors array,” Sensors Actuators B Chemical, vol. 69, no. 3, Oct. 2000, pp. 230-235.

[8] A. P. F. Turner and N. Magan, “Electronic noses and disease diagnostics,” Nature Reviews Microbiology 2, no. 2, Feb. 2004, pp. 161-166.

[9] M. Trincavelli, S. Coradeschi, A. Loutfi, B. Sderquist, P. Thunberg, “Direct identification of bacteria in blood culture samples using an electronic nose,” IEEE Trans. Biomed. Engineering, vol. 57, no. 12, Dec. 2010, pp. 2884-2890.

[10] R. E. Baby, M. Cabezas, and E. N. Walse de Reca, “Electronic nose: a useful tool for monitoring environmental contamination,” Sensors Actuators B Chemical, vol. 69, no. 3, Oct. 2000, pp. 214-218.

[11] M. Penza and G. Cassano, “Classification of food, beverages and perfumes by WO3 thin-film sensors array and pattern recognition techniques,” Sensors Actuators B Chemical, vol. 73, no. 1, Feb. 2001, pp. 76-87.

[12] M. Holmberg, F. Winquist, I. Lundstrm, J. W. Gardner, E. L. Hines, “Identification of paper quality using a hybrid electronic nose,” Sensors Actuators B Chemical, vol. 27, no. 1, Jun. 1995, pp. 246-249.

[13] T. W. Margrie and A. T. Schaefer, “Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system,” J. Physiology, vol. 546, no. 2, Nov. 2002, pp. 363-374.

[14] H. T. Chen, K. T. Ng, A. Bermak, M. K. Law, D. Martinez, “Spike latency coding in biologically inspired microelectronic nose,” IEEE Trans. Biomed. Circuits and Systems, vol. 5, no. 2, Apr. 2011, pp. 160-168.

[15] M. Hassan, S. Brahim Belhaouari, A. Bermak, “Probabilistic rank score coding: A robust rank-order based classifier for electronic nose applications,” IEEE Sensors Journal, vol. 15, no. 7, Jul. 2015, pp. 3934-3946.

[16] M. Hassan and A. Bermak, “Threshold detection of carcinogenic odor of formaldehyde with wireless electronic nose,” in IEEE Sensors, Valencia, November 2014, pp. 1368-1371.

[17] J. A. Yamani, F. Boussaid, A. Bermak, D. Martinez, “Glomerular latency coding in artificial olfaction,” Front. Neuroengineering, vol. 4, art. 18, Jan. 2012, pp. 1-9.

[18] E. Alpaydin, Introduction to Machine Learning, MIT Press, 2010.

[19] D. C. Knill and A. Pouget, “The Bayesian brain: the role of uncertainty in neural coding and computation,” Trends in Neuroscience, vol. 27, no. 12, Dec. 2004, pp. 712-719.

 ISBN: 978-1-61208-426-8