

Estimation of the Number of Volatile Compounds in Simple Mixtures [†]

Olivier Hotel ^{*}, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone and Samuel Saada

CEA, LIST, 91191 Gif-sur-Yvette, France; jean-philippe.poli@cea.fr (J.-P.P.); christine.mer@cea.fr (C.M.-C.); emmanuel.scorsone@cea.fr (E.S.); samuel.saada@cea.fr (S.S.)

^{*} Correspondence: olivier.hotel@cea.fr

[†] Presented at the EuroSensors 2017 Conference, Paris, France, 3–6 September 2017.

Published: 11 August 2017

Abstract: In this paper, we address the problem of simple mixtures identification using a SAW sensor based e-nose. First, we propose a linearized form of the SAW sensor's response to mixtures. This equation is then used to set up a linear least squares problem whose residuals are suitable to identify the number of compounds present in a mixture thanks to a supervised learning algorithm.

Keywords: SAW sensor; mixture identification; pattern recognition

1. Introduction

In this paper, we propose an approach to tackle the problem of simple mixture identification using an array of SAW sensors. These sensors are based on the propagation of mechanical waves produced by piezoelectric materials along a layer coated with chemically interactive materials. Volatile compounds are absorbed onto the surface of the sensitive material, changing its properties and yielding to a measurable frequency shift. In [1], authors established that the frequency shift is the superposition of a mass loading effect and of a viscoelastic contribution. Both can be modelled by first order linear differential equations $\tau_i \frac{\partial F_i}{\partial t} + F_i = K_i c(t), i \in \{m, v\}$ where F_m and F_v are respectively the frequency shift due to the mass loading effect and to the viscoelastic contribution, τ_m, τ_v, K_m and K_v are the time constants and the gains of the contribution and $c(t)$ is the concentration profile which is not known. The total frequency shift is then given by $F = F_m + F_v$. One can show, by discretizing the differential equations and solving the recurrence equations, that the expression of $F[n]$ can be written as $F[n] = A_m T_m^n + A_v T_v^n + P[n]$ where $P[n]$ is the sum of the particular solution of the discretized equation for the two contributions. In this paper, we assume that the response of the sensor to a mixture is a weighted sum of the responses to each individual compounds:

$$F[n] = \sum_{i=1}^{N_g} \gamma_i F_i[n] = \sum_{i=1}^{N_g} \gamma_i A_{m,i} T_{m,i}^n + \gamma_i A_{v,i} T_{v,i}^n + \gamma_i P_i[n] \quad (1)$$

This assumption can be verified experimentally. The selected e-nose was composed of 6 functionalized diamond coated SAW sensors. The sensors were exposed to $NH_3, SO_2, H_2S, CH_3OH, C_7H_8$, to the binary mixtures $SO_2 + C_7H_8, NH_3 + CH_3OH, NH_3 + C_7H_8, H_2S + NH_3, H_2S + CH_3OH$ and to the ternary mixtures $H_2S + NH_3 + SO_2, H_2S + NH_3 + CH_3OH, H_2S + C_7H_8 + SO_2$. Several cycles exposition—purge were done. The assumption was tested by solving a least squares optimization problem to estimate the coefficients γ , and by measuring the relative error between the true mixture signals and the one reconstructed from the pure signals. Tables 1 and 2 show the average relative error.

Table 1. Binary mixture reconstruction error.

Mixture	$SO_2 + C_7H_8$	$NH_3 + CH_3OH$	$H_3 + C_7H_8$	$H_2S + NH_3$	$H_2S + CH_3OH$
Relative error	23.01%	10.89%	19.9%	11.47%	12.84%

Table 2. Ternary mixture reconstruction error.

Mixture	$H_2S + NH_3 + SO_2$	$H_2S + NH_3 + CH_3OH$	$H_2S + C_7H_8 + SO_2$
Relative error	24.6%	55.62%	37.76%

These tables show that the assumption is plausible for the binary mixtures since the average error is near the variance of the successive experiments and that the assumption does not hold for the ternary mixtures. However, this is still informative since the objective consists in estimating the number N_g of compounds in simple mixture.

2. Estimation of the Number of Simple Mixture's Components

2.1. Notation and Definitions

In the rest of this paper, $P^d[\cdot]$ denotes a polynomial of degree d , $P_k(E)$ the set of subsets of E of cardinality equal to k , S the cumulative sum and $S^k_a = \begin{cases} a & \text{if } k = 0 \\ S(S^{k-1}_a) & \text{otherwise} \end{cases}$ is the k^{th} iterated cumulative sum of the sequence a .

2.2. Linearisation of the Sensor's Response

Prior to propose a linear form for the sensor's response, we remind that a direct consequence of the Faulhaber's formula allows to conclude that the k^{th} iterated cumulative sum of a sequence whose general term is a polynomial of degree d is a polynomial of degree $d + k$. Moreover, it can be easily be prooved by induction that the k^{th} iterated cumulative sum of a sequence whose general term is a sum of geometric series $g[n] = \sum_i a_i r_i^n$ can be written in the form $S^k_g = \sum_i a_i \left(\frac{r_i}{r_i - 1}\right)^k r_i^n + P^{k-1}[n]$.

Equation (1) can be rewritten $F[n] = \sum_{i=1}^{2N_g} A_i T_i^n + P[n]$. For the sake of clarity, we note \mathfrak{T} the set $[1 \cdots 2N_g]$. Let's compute $F[n] + \sum_{i=1}^{2N_g} (-1)^i S^i_F[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}$. The previously described results and the fact that the iterated cummulative sum operator is linear yield to:

$$F[n] + \sum_{i=1}^{2N_g} (-1)^i S^i_F[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} = F[n] + \sum_{i=1}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1}\right)^i T_k^n \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} + \sum_{i=1}^{2N_g} (-1)^i P^{i-1}[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} + \sum_{i=1}^{2N_g} (-1)^i S^i_{S_p}[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}.$$

First, one can notice that the term $\sum_{i=1}^{2N_g} (-1)^i P^{i-1}[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}$ is a weighted sum of polynomials of degree up to $2N_g - 1$ so it is a polynomial P^{2N_g-1} of degree $2N_g - 1$. Secondly, the term $\sum_{i=1}^{2N_g} (-1)^i S^i_{S_p}[n] \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}$ is a sequence which will be considered as an unknown in the next section. For the sake of clarity, this term is denoted as $P_s[n]$ in the rest of this article. Finally, one can remark that the term $\sum_{i=1}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1}\right)^i T_k^n \sum_{I \in P_i(\mathfrak{T}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}$ is equal to 0. Indeed, this term can be simplified by remarking that the sum indexed by the set I can be split into two parts:

$$\begin{aligned}
 & A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &= A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_{i-1}(\mathfrak{I} \setminus \{k\}) \setminus \emptyset} \frac{T_k - 1}{T_k} \prod_{j \in I} \frac{T_j - 1}{T_j} + A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I} \setminus \{k\}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &= A_k \left(\frac{T_k}{T_k - 1} \right)^{i-1} T_k^n \sum_{I \in P_i(\mathfrak{I} \setminus \{k\}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} + A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I} \setminus \{k\}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}.
 \end{aligned}$$

Two specific cases should be noticed: $i = 1 \rightarrow P_{i-1}(\mathfrak{I} \setminus \{k\}) \setminus \emptyset = \emptyset$ and $i = 2N_g \rightarrow P_i(\mathfrak{I} \setminus \{k\}) \setminus \emptyset = \emptyset$. Consequently, we have:

$$\begin{aligned}
 & \sum_{i=1}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &= - \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^1 T_k^n \sum_{I \in P_1(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &+ \sum_{i=2}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^{i-1} T_k^n \sum_{I \in P_{i-1}(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} + \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &+ \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^{2N_g-1} T_k^n \sum_{I \in P_{2N_g-1}(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}
 \end{aligned}$$

Moreover, we have:

$$\begin{aligned}
 & \sum_{i=2}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^{i-1} T_k^n \sum_{I \in P_{i-1}(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &+ \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} = \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^1 T_k^n \sum_{I \in P_1(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} \\
 &- \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^{2N_g-1} T_k^n \sum_{I \in P_{2N_g-1}(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}
 \end{aligned}$$

since the series are telescoping. Hence, we have $\sum_{i=1}^{2N_g} (-1)^i \sum_{k=1}^{2N_g} A_k \left(\frac{T_k}{T_k - 1} \right)^i T_k^n \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j} = 0$.

Hence, by posing $\alpha_i = -(-1)^i \sum_{I \in P_i(\mathfrak{I}) \setminus \emptyset} \prod_{j \in I} \frac{T_j - 1}{T_j}$, Equation (1) can be rewritten in the linear form:

$$F[n] = \sum_{i=1}^{2N_g} \alpha_i S_F^i[n] + P^{2N_g-1}[n] + P_s[n] \quad (2)$$

2.3. Optimization Problem Formulation

In this section, we assume that N samples $F[0] \dots F[N-1]$ were digitalized. We define the vector F and C of respective length N and $4N_g + N$: $F = [F[0] \dots F[N-1]]^T$, $C = [\alpha_1 \dots \alpha_{2N_g} p_0 \dots p_{2N_g-1} P_s[0] \dots P_s[N-1]]^T$. And the $(N) \times (2N_g)$ matrices S and P : $S = \begin{bmatrix} S_F^1[0] & \dots & S_F^{2N_g}[0] \\ \vdots & & \vdots \\ S_F^1[N-1] & \dots & S_F^{2N_g}[N-1] \end{bmatrix}$, $P = \begin{bmatrix} 1 & 0^1 & \dots & (N-1)^{2N_g-1} \\ \vdots & \vdots & & \vdots \\ 1 & (N-1)^1 & \dots & (N-1)^{2N_g-1} \end{bmatrix}$ and we form the matrix $X = [S \ P \ I_N]$. With these definitions, Equation (2) can be rewritten as $F = XC$. As the number of variables is greater than the number of equations ($4N_g + N$ variables vs. N equations), we should add a regularization term to avoid overfitting. We define the matrix $\Gamma = \begin{bmatrix} I_{4N_g} & 0_{4N_g, N} \\ 0_{N-2, 4N_g} & D \end{bmatrix}$ where D is the second order differentiation matrix. The identity matrix prevents the coefficients α and p to grow unbounded, and the matrix D smooths the unknown sequence P_s . The parameters can be estimated by solving the optimization problem $\hat{C} = \argmin \|F - XC\|^2 + \lambda \| \Gamma C \|^2$ whose solution is

$\hat{C} = \begin{bmatrix} X \\ \sqrt{\lambda} \Gamma \end{bmatrix}^+ \begin{bmatrix} F \\ 0 \end{bmatrix}$ [2]. The residuals are defined as $R = F - X\hat{C}$ and the mean squared error (MSE) as $MSE = \frac{1}{N} R^T R$.

2.4. Experimental Results

In this section, we propose an approach, based on supervised learning, to estimate the number of compounds in simple mixtures. In particular, we compare the performances of three different sets of features. The first set is composed of the steady state amplitude of the response of each sensors whereas the second set is composed of the RMS error obtained from the least squares optimization problem for $N_g = 1, 2, 3$ and for each sensor. The third set is the concatenation of these two sets. Table 3 gives the results obtained over a 5-fold cross validation process using the LMNN algorithm [3]. It exhibits that the proposed features are only outperformed by the steady state when $N_g = 3$. It also shows the interest of using both features since in this case the performances are increased.

Table 3. Performances obtained using the steady state amplitude (1), the MSE errors (2) and the concatenation of the two feature spaces (3).

		Predicted N_g			Predicted N_g			Predicted N_g				
		1	2	3	1	2	3	1	2	3		
Actual N_g	1	90.4%	8.1%	1.5%	1	93.1%	6.9%	0.0%	1	96.3%	3.2%	0.5%
	2	25.6%	72.0%	2.6%	2	19.7%	76.7%	3.6%	2	6.7%	92.8%	0.4%
	3	9.0%	9.0%	82.0%	3	8.1%	11.1%	80.8%	3	3.0%	3.0%	94.0%
		(1)			(2)			(3)				

3. Conclusions

In this paper, we established that the response of SAW sensors to a simple mixture of volatile compounds can be linearized thank to iterated cumulative sums. This linear equation allows the formulation of a regularized linear least squares problem whose residuals, in conjunction with the steady state amplitudes of the signals and supervised learning algorithms, are suitable to estimate the number of compounds in a mixture (the estimation accuracy is higher than 90%).

Acknowledgments: This research was supported by the French trans-governmental CBRN-E R & D program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Raj, V.; Nimal, A.; Parmar, Y.; Sharma, M.; Gupta, V. Investigations on the origin of mass and elastic loading in the response of ZnO SAW ammonia sensor. *Sens. Actuator B* **2012**, *166*, 576–585.
2. Golub, G.; Loan, C. *Matrix Computation*; The Johns Hopkins University Press: Baltimore, MD, USA, 2013.
3. Weinberger, K.; Blitzer, J.; Saul, L. Distance metric learning for large margin nearest neighbor classification. *Adv. Neural Inf. Process. Syst.* **2006**, *18*, 1473–1480.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).