Available online at www.sciencedirect.com
ScienceDirect
Sensors and Actuators B xxx (2007) xxx–xxx
SENSORS AND ACTUATORS B CHEMICAL
www.elsevier.com/locate/snb

Mix-to-mimic odor synthesis for electronic noses

Liran Carmel [1], David Harel *

Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 30 August 2006; received in revised form 5 March 2007; accepted 9 March 2007

Abstract

Arrays of chemical sensors, known as electronic noses, yield a unique pattern for a given mixture of odors. Recently, there has been increasing interest in trying to mix odors such as to generate a desired response in the electronic nose. For the time being, this intriguing problem had been tackled only experimentally with the aid of specific apparatus. Here, we present an algorithmic solution to the problem. We demonstrate the algorithm on data that includes mixtures of up to five ingredients.

Keywords: Odor communication; Sniffer; Whiffer; Within-sniffer mix-to-mimic algorithm; Electronic nose

1. Introduction

Sights and sounds were long ago proved amenable to digital manipulation: storage, compression, and re-formation at the end user's site. Olfaction still lags far behind. While fragrances are developed and manufactured in laboratories worldwide, their digitization is underdeveloped, and their part in modern multimedia is limited to a paucity of anecdotal applications.

In recent years, we have worked on odor digitization, particularly on *odor communication*, defined in Harel et al. [1] A *mix-to-mimic* algorithm (M2M)[2] instructs an output device, the *whiffer*, to release an imitation of an odorant[3] read in by a remote input device, the *sniffer*, which is to digitize smells in a way that preserves relevant sensory information. The whiffer contains a fixed set of *palette odorants*, a technology to mix them accurately and means to release them in precise quantities and with precise timing. The M2M algorithm instructs the whiffer as to the ratios in which to mix the palette odorants.

While current technology can produce sniffers and whiffers, knowledge of the sense of smell does not yet allow for full development of the M2M algorithm. In Harel et al. [1] we suggested that this should be done by constructing three increasingly complex sub-algorithms, the third of which constitutes the full algorithm:

- *Within-sniffer mix-to-mimic* (WSM2M), or "fooling a sniffer". Given a sniffer-generated digital fingerprint (pattern) of an odorant, compute the palette mixture whose pattern, as generated by the same sniffer, best resembles the original pattern. This is a sniffer-limited version of the M2M algorithm.
- *Between-sniffers mix-to-mimic* (BSM2M), or "fooling a different sniffer". Given the fingerprint of an odorant generated by sniffer S_1, compute the palette mixture whose pattern, as generated by sniffer S_2, best resembles the pattern S_2 would have generated for the original odorant. This requires mapping one sniffer's fingerprints to another's, which could be very complicated when the sniffers are of different nature.
- *Full mix-to-mimic* (M2M), or "fooling the human brain". Given the sniffer-generated fingerprint of an odorant, compute the palette mixture which, sniffed by a human, generates a sensation that best resembles the one perceived by sniffing the input odorant. This requires translating digital fingerprints to perception-related "patterns". We proposed [1] that this will be done using human panels and working within a psychophysical perception space.

This paper describes the work we have done in constructing a complete WSM2M algorithm designed for an electronic nose (eNose) as the sniffer.

* Corresponding author. Tel.: +972 8 9344 050; fax: +972 8 9344 122.
E-mail address: dharel@weizmann.ac.il (D. Harel).
[1] Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.
[2] We uses the acronym MTM in Harel et al. [1], but these days prefer M2M.
[3] We use *odorant* to refer to any distinct odor-stimulus, whether pure compound or mixture.

0925-4005/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.snb.2007.03.022

Finding a mixture that elicits a desired response has already raised interest in the eNose community. The best solution to date [2,3] involves a special apparatus, called an *active odor system*, used to iteratively adapt the concentration of the palette odorants. Although this seems to work well when given prior information on the mixture ingredients, it requires a measurement at each iteration, which limits its utilization. Our solution is purely computational, requiring only an input pattern and some pre-measured properties of the palette odorant.

2. Electronic noses

The desired properties of an ideal sniffer are discussed in detail in Harel et al. [1]. These can be roughly summarized as the ability to translate chemical information into numbers in a sufficiently discriminatory fashion, and in such a way that the fingerprints will show some correlation with the human smell perception. As can be expected, existing sniffers are not ideal, and we have chosen to work with devices known collectively as *electronic noses* (eNose) that seem to have a set of particularly appealing properties.

An eNose is an analytic device that hosts a multiplicity of non-specific chemical sensors that interact with a broad range of chemicals with varying strengths, eliciting unique response patterns. The first eNoses were developed in the early 1980s [4], and since then many different types have been designed, employing a variety of sensor technologies.

The fact that the biological smelling system also relies on an array of non-specific receptors [5] gives hope that we may be able to find significant relationships between the biological nose and its electronic counterpart. Indeed, in Harel et al. [1] we present evidence of the existence of such relationships. Additional support is discussed briefly in Section 6. Moreover, elsewhere we have shown that a model originally suggested to explain odor information processing in the brain [6] can be easily adapted to eNoses, yielding an algorithm that can recognize odorants and estimate their concentration [7].

Electronic noses have been designed, first and foremost, to deal with the classification problem, which is the task of determining the identity of incoming stimuli. Indeed, eNoses seem to fulfill their designation pretty successfully in a wide range of applications. For example, they are used for medical diagnostics [8] for environmental control [9] and for quality assessment of food products [10,11].

Dealing with mixtures, e.g., revealing their chemical makeup was never intended to be within the capabilities of eNoses. For most stimuli, there is nothing in the pattern elicited by a mixture that can be used to discern it from a pattern elicited by a pure compound. Exactly the same methodology should be used in order to train an eNose to discriminate between methanol and ethanol [12] or to discriminate between different types of olive oils [13]. Traditionally, indeed, eNoses are seldom used for mixture analysis. One typical exception is the case when two mixtures are formed by taking different ratios of the same ingredients, thus producing mixture-dependent response patterns. This can be used to design an algorithm that estimates mixing ratios in a mixture whose ingredients are known [14].

An odor communication system [1] is not required for mixture composition analysis either. However, such a system must be able to accomplish the inverse task—finding an appropriate mixture from within given ingredients that yields some desired signal. This, of course, requires understanding how patterns of mixtures are related to those of the individual ingredients.

In order to formally express such relations, we introduce some notation. Let $(o; c)$ stand for a pure chemical o in concentration c. When measured by an eNose, this translates into a list of m features, thus giving rise to an m-dimensional *response vector*, $r(o; c)$. Each feature changes in a way dependent on the concentration, a function that we shall dub the *response curve*. The m response curves of an odorant o completely characterize the eNoses behavior with respect to this odorant. That is, given the odorant concentration, the response vector can be predicted straightforwardly from these curves.

Let $r(o_1, \ldots, o_n; c_1, \ldots, c_n)$ be the eNose response to the mixture of the pure chemicals $o_1, \ldots, o_n$ in concentrations $c_1, \ldots, c_n$, respectively. In general, even if we know all the response curves of all the pure chemicals, it is not straightforward to tell what $r(o_1, \ldots, o_n; c_1, \ldots, c_n)$ would be, since the response curve of a given stimulus is modified in the presence of other stimuli. In previous work [15] we experimentally examined the relations between the response vector of the mixture and the response vectors of the pure ingredients. We were able to show that the *linear law of mixtures*:

$$r(o_1, \ldots, o_n; c_1, \ldots, c_n) = \alpha_1 r(o_1; c_1) + \cdots + \alpha_n r(o_n; c_n), \tag{1}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are constants that we call the *mixing coefficients*, describes these relations fairly precisely, with the deviation between the measured mixture response vector and the predicted one rarely exceeding a few percent. The mixing coefficients are, obviously, palette-dependent, and should be determined by a series of preliminary experiments that have to be carried out once for each given palette. In Carmel et al. [15] we describe these experiments, and show how to use them for computing the mixing coefficients. Hereinafter, we shall assume that the mixing coefficients have already been computed for each palette that we use.

3. The within-sniffer mix-to-mimic algorithm

Let the palette of a whiffer consist of n *palette odorants* $t_1, \ldots, t_n$. Given any odorant $(o; c)$, we are interested in finding the *mixing vector* $v = (v_1, \ldots, v_n)^{\mathrm{T}}$ that is the solution of

$$v = \begin{cases} \underset{v}{\text{argmin}} & \|r(t_1, \ldots, t_n; v_1, \ldots, v_n) - r(o; c)\| \\ \text{such that} & v_i \geq 0 \ \ \forall i. \end{cases} \tag{2}$$

For mathematical tractability we shall hereinafter assume the L_2 norm.

Had all the response curves been linear, this problem is readily solved. To this end, let the response curve of the jth feature to the ith palette odorant be $r_j(t_i; v_i) = \beta_{ji} v_i$. Substituting this in the linear law of mixtures (1) and using (2), v is found to be a

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

solution of

$$v = \begin{cases} \underset{v}{\text{argmin}} & ||\mathcal{A}v - r(o;c)|| \\ \text{such that} & v_i \geq 0 \ \ \forall i, \end{cases} \tag{3}$$

where $\mathcal{A}$ is the $m \times n$ matrix $\mathcal{A}_{ij} = \alpha_j \beta_{ij}$. This is a well-studied optimization problem, known as the non-negative least squares [16].

In reality, however, the response curves are rarely linear [7]. Yet, we can still use the linear paradigm to iteratively solve the non-linear problem. Assume, then, that the response curves $r_j(t_i; v_i)$ are non-linear, and that we have achieved, in the pth iteration, an approximation v^p of the mixing vector. Let us define the linear response curves $r_j^p(t_i; v_i)$ as the first order Taylor expansion of $r_j^p(t_i; v_i)$ around v_i^p, i.e.,

$$r_j^p(t_i; v_i) = r_j(t_i; v_i^p) + \left(\frac{\mathrm{d}r_j(t_i; v_i)}{\mathrm{d}v_i} \right)_{v_i^p} (v_i - v_i^p).$$

Using the notation $\beta_{ji}^p = (\mathrm{d}r_j(t_i; v_i)/\mathrm{d}v_i)_{v_i^p}$, the response of the palette mixture can consequently be approximated by

$$\begin{aligned} & r_j(t_1, \ldots, t_n; v_1, \ldots, v_n) \\ & \cong \alpha_1 r_j^p(t_1; v_1) + \cdots + \alpha_n r_j^p(t_n; v_n) \\ & = \alpha_1 (r_j(t_1; v_1^p) + \beta_{j1}^p (v_1 - v_1^p)) + \cdots \\ & + \alpha_n (r_j(t_n; v_n^p) + \beta_{jn}^p (v_n - v_n^p)). \end{aligned} \tag{4}$$

Let us now define the *target vector* as

$$\tau^p = r(o;c) - [\alpha_1 r(t_1; v_1^p) + \cdots + \alpha_n r(t_n; v_n^p)], \tag{5}$$

which is really the error that one makes in reproducing $r(o; c)$ by using the mixing vector v^p. Then, the optimization problem reduces to $\mathrm{d}v^p = \text{argmin}_{\mathrm{d}v} ||\mathcal{A}^p \, \mathrm{d}v - \tau^p||$, where $\mathcal{A}_{ij}^p = \alpha_j \beta_{ij}^p$ and $\mathrm{d}v = v - v^p$. The idea is, of course, to find a correction to v such as to reduce the previous error. Such a scheme requires using small correction in each iteration, leading to the following minimization problem:

$$\mathrm{d}v^p = \underset{\mathrm{d}v}{\text{argmin}} ||\mathcal{A}^p \, \mathrm{d}v - \gamma \tau^p||, \tag{6}$$

where $0 \leq \gamma \leq 1$ is a parameter of the algorithm. Having found $\mathrm{d}v^p$, we find the improved approximation $v_i^{p+1} = \max(v_i^p + \mathrm{d}v_i^p, 0)$. Note that (6) is solved using ordinary least squares and not non-negative least squares as in Eq. (3). The updated target vector is then

$$\tau^{p+1} = r(o;c) - [\alpha_1 r(t_1; v_1^{p+1}) + \cdots + \alpha_n r(t_n; v_n^{p+1})].$$

The actual value of γ is not very important, since if in a certain iteration we are not able to reduce the error, that is, if $||\tau^{p+1}|| > ||\tau^p||$, we repeatedly substitute $\gamma \leftarrow \gamma/2$ and recompute the iteration until the error is reduced.

The linear solution is also used to initialize the iterative process. To better understand how this is done, we elaborate on how the response curves are computed [15]. For each palette odorant i, we take a series of K known concentrations, $v_i^1, \ldots, v_i^K$, and measure for each the response curves $r_j(v_i^k)$, $j = 1, \ldots, m$, $k = 1, \ldots, K$. The non-linear response curves are interpolated from the measured points using cubic spline interpolation. But, for purposes of initialization, we assume a linear response curve, $r_j(t_i; v_i) = \beta_j^0(i) v_i$, and find the coefficients $\beta_j^0(i)$ as the solution of the univariate least squares problem:

$$\beta^0(i) = \underset{\beta}{\text{argmin}} \sum_{k=1}^{K} (\beta v_i^k - r_j(t_i; v_i^k))^2. \tag{7}$$

The initial approximation v^1 of the mixing vector is just the non-negative least squares solution of Eq. (3), in which we use $\beta_{ji} = \beta_j^0(i)$.

The full non-linear WSM2M algorithm is summarized in Algorithm 1. Note that convergence to the global minimum cannot be guaranteed. However, our experience is that the non-linear response curves are approximated pretty well by their linear counterparts, resulting in rather smooth target functions during the minimization process. Hence, in practice, the global minimum is often found.

Algorithm 1. The full (non-linear) WSM2M algorithm

Function optimize $(\{r(t_i; v_i)\}_{i=1}^n, \alpha, r(o;c), \varepsilon, \gamma_0)$
% $\{r(t_i; v_i)\}_{i=1}^n$ is the set of n response curves, where
  $r(t_i; v_i) = (r_1(t_i; v_i), \ldots, r_n(t_n; v_n))^{\mathrm{T}}$
% α is the vector of mixing coefficients
% $r(o;c)$ is the measured response of the incoming odorant
% ε is the tolerance
% γ_0 is the step size

% initialization
for $i = 1$ to n
  $\beta^0(i) \leftarrow$ solution of (7)
  for $j = 1$ to m
    $\mathcal{A}_{ji} \leftarrow \alpha_i \beta_{ji}^0$
  end for
end for
$v^1 \leftarrow$ non-negative least squares solution of $||\mathcal{A}v - r(o;c)||$.

% iterations
$p \leftarrow 0$
repeat
  $p \leftarrow p + 1$
  $\gamma \leftarrow \gamma_0$
  % compute linear coefficients and target vector
  $\tau \leftarrow r(0;c)$
  for $i = 1$ to n
    for $j = 1$ to m
      $\beta_{ji}^p \leftarrow (\mathrm{d}r_j(t_i; v_i)/\mathrm{d}v_i)_{v_i^p}$
      $A_{ji}^p \leftarrow \alpha_i \beta_{ji}^p$
      $r_j^p(t_i; v_i) \leftarrow r_j(t_i; v_i^p) + \beta_{ji}^p (v_i - v_i^p)$
    end for
    $\tau^p \leftarrow \tau^p - \alpha_i r(t_i; v_i^p)$
  end for
  % find least squares solution
  repeat
    $\mathrm{d}v^p \leftarrow \text{argmin}_{\mathrm{d}v} ||A^p \, \mathrm{d}v - \gamma \tau^p||$
    $v^{p+1} \leftarrow \max(v^p + \mathrm{d}v^p, 0)$
    $\tau^{p+1} \leftarrow r(o;c) - \alpha_1 r(t_1; v_1^{p+1}) - \cdots - \alpha_n r(t_n; v_n^{p+1})$
    $\gamma \leftarrow 1/2\gamma$
  until $||\tau^{p+1}|| \leq ||\tau^p||$
until $(||\tau^p|| - ||\tau^{p+1}||)/||\tau^p|| < \varepsilon$

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

Table 1
Pure chemicals and their concentrations (data taken from Carmel et al. [15])

Chemical (abbreviation)	Concentrations measured (molar fraction)					
1-Methylpyrrole (M)	0.0908	0.1665	0.2306	0.2855	0.3331	0.3747
1-Propanol (P)	0.1055	0.1909	0.2614	0.3206	0.3710	0.4144
2,3-Butanedione (B)	0.0918	0.1316	0.1681	0.2016		
2,6-Dimethylpyridine (D)	0.0711	0.1328	0.1867	0.2344	0.2768	0.3147
2-Methyl-2-pentenal (MP)	0.0721	0.1345	0.1890	0.2371	0.2797	0.3179
4-Methylanisole (MA)	0.0657	0.1233	0.1742	0.2195	0.2601	0.2967
Amyl formate (A)	0.0632	0.0919	0.1189	0.1443	0.1683	0.1910
Butyl butyrate (BB)	0.0983	0.1406	0.1791	0.2142	0.2465	
Isoamyl formate (I)	0.0633	0.0920	0.1190	0.1445	0.1685	0.1912
Toluene (T)	0.0770	0.1112	0.1430	0.1726	0.2002	0.2260

Molar fractions are measured in PEG-400 solution. The abbreviated names are used extensively throughout the paper.

4. Materials and methods

We have been using the MosesII eNose [17], hosting eight quartz-microbalance (QMB) sensors, which are well known for the fact that their response curves deviate not-too-strongly from linear [18,19]. We define the response of a sensor in the traditional way by taking the difference between the maximum of the signal and its baseline. This eNose is designed for precise laboratory work, obtaining its input from an accompanying headspace sampler (HP7694). The sampler prepares all samples at the same temperature and pressure, and injects them at the same rate into the eNose. In all our experiments, the samples were inserted into the headspace sampler in 20-ml vials. Then, the headspace sampler heated them to 40 °C and injected the headspace content into the eNose in a flow of 25 ml/min. The injection lasted for 20 s, and was followed by a 15 min purging stage using synthetic air. Each stimulus was measured in batches, with a single batch containing several successive measurements.

We collected two different datasets, already used to infer the laws of mixture and described in Carmel et al. [15]:

- The *pure odorants dataset* was constructed from 10 pure chemicals, each measured in six different concentrations[4]; see Table 1. The concentrations were chosen in order for all the chemicals to have comparable ranges of response. Each sample was diluted in polyethylene glycol 400 (PEG-400), and the concentrations were measured in molar fractions. A chemical in a certain concentration was measured in batches of at least four successive measurements. In total, this dataset consists of 269 measurements.
- The *mixtures dataset* was used to test the WSM2M algorithm. It consisted of binary, trinary, quaternary and quinary mixtures of the above palette odorants, as listed in Supplementary Tables S1 and S2. Each of the mixtures was diluted in PEG-400 to obtain six different total concentrations for the same mixing ratios (not shown in tables). Each specific mixture dilution was measured in batches of about seven successive measurements. All in all, there were 27 binary mixtures (1095 measurements), 10 trinary mixtures (409 measurements), 11 quaternary mixtures (452 measurements), and 1 quinary mixture (42 measurements). In total, there were 49 different mixtures (each in six different total concentrations) and 1998 measurements.

We measured concentrations as molar fractions in PEG-400 solution. However, for each compound this number is proportional to its headspace concentration, as explained in Carmel et al. [15]. These data were used in our previous work to compute the mixing coefficients α for each mixture of palette odorants.

5. Results

5.1. Definition of prediction error

We have tested our algorithm by feeding a known mixture of palette odorants $v^0 = (v_1^0, \ldots, v_n^0)^{\mathrm{T}}$ into our eNose, obtaining the response vector r. Then, we used the WSM2M algorithm to compute a mixing vector v that would best reconstruct r. The feasibility of our algorithm can be appraised by comparing v^0 with $\hat{v}$.

Such a comparison, however, must be carried out with caution, as a naive straightforward approach can be misleading. This is exemplified in Fig. 1a, where actual concentrations are compared to the predicted ones for each palette odorant in a quaternary mixture. At first glance, it looks that the algorithm does a good job with respect to 1-methylpyrrole and amyl formate, a moderate job for 1-propanol, and a terrible job for 2-methyl-2-pentenal. Such a conclusion is highly misleading. For this particular palette, the mixing coefficients are $\alpha = (1.23, 0.3, -0.003, 1.8)^{\mathrm{T}}$, suggesting that 2-methyl-2-pentenal has a negligible effect on the response vector of the mixture, 1-propanol has a moderate effect, and 1-methylpyrrole and amyl formate have the largest effect.

We therefore define the prediction error of palette odorant concentrations in a way that accommodates the mixing coefficients. Keeping all concentrations fixed except for that of the ith palette odorant, we get from (1):

$$\Delta\hat{v}_i = \frac{||\Delta r(t_1, \ldots, t_n; \hat{v}_1, \ldots, \hat{v}_n)||}{|\alpha_i|||\mathrm{d}r(t_i; \hat{v}_i)/\mathrm{d}\hat{v}_i||}. \quad (8)$$

[4] Except for 2,3-butanedione, which was measured only in four concentrations, and butyl butyrate, which was measured in five concentrations.

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

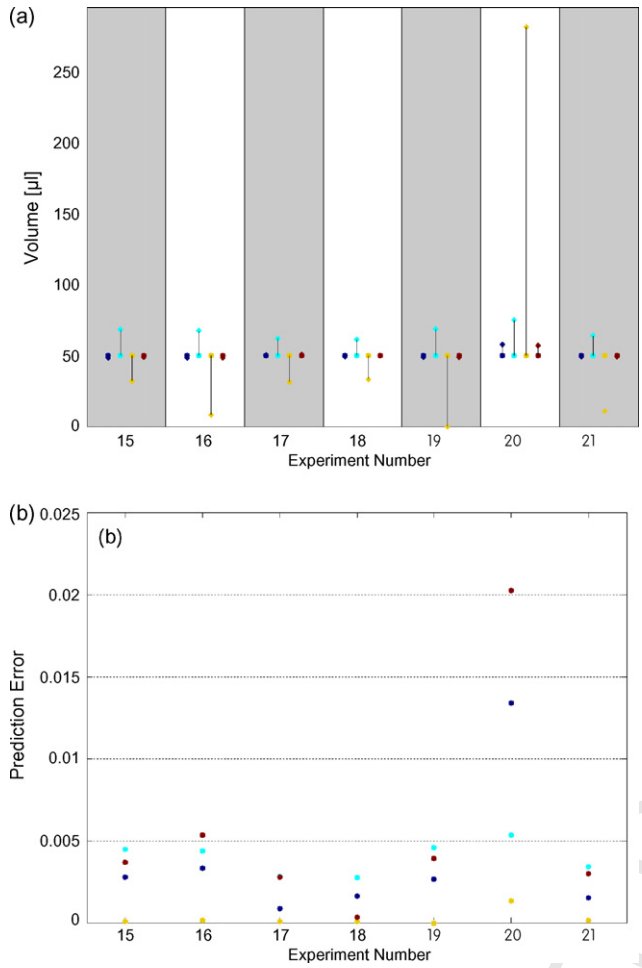

Fig. 1. A portion from a series of experiments with a quaternary palette. The palette consists of 1-methylpyrrole, 1-propanol, 2-methyl-2-pentenal and amyl formate, denoted by blue, cyan, yellow and red, respectively. (a) Actual concentrations taken (circles) vs. predicted concentrations (diamonds); (b) prediction errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

This equation estimates the uncertainty in $\hat{v}_i$ given the uncertainty in the response vector $r(t_1, \ldots, t_n; \hat{v}_1, \ldots, \hat{v}_n)$. Here, α_i is known, and $||\mathrm{d}r(t_i; \hat{v}_i)/\mathrm{d}\hat{v}_i||$ is readily computed from the response curves of the pure palette odorants. The standard deviation of a general response vector in our eNose is estimated to be in the range of 5–10% of the average response.[5] So, to estimate the uncertainty $\Delta\hat{v}_i$, we have taken

$$\Delta r(t_1, \ldots, t_n; \hat{v}_1, \ldots, \hat{v}_n) \cong 0.075 r(t_1, \ldots, t_n; \hat{v}_1, \ldots, \hat{v}_n).$$

To account for the uncertainty in $\hat{v}_i$, we define the "distance" between $\hat{v}_i$ and v_i^0 to be the Mahalanobis distance $|\hat{v}_i - v_i^0|/\Delta\hat{v}_i$. Consequently, we define the prediction error for the ith palette odorant as

$$\varepsilon_i = \frac{2|\hat{v}_i - v_i^0|}{\Delta\hat{v}_i(\hat{v}_i + v_i^0)}. \quad (9)$$

[5] Private communication with MoTech GmbH, which is at the time of writing part of AppliedSensors GmbH; see http://www.appliedsensor.com.

Fig. 1b plots the prediction errors of the same data shown in Fig. 1a. Now, the prediction error of 2-methyl-2-pentenal is nearly zero despite the relatively large values of $|\hat{v}_i - v_i^0|$. We see that in most of the experiments the prediction errors are below 0.6%.

5.2. Mixture reconstruction

In the first set of experiments, we used many different palettes (7 binary, 10 trinary, 7 quaternary and 1 quinary). For each palette, we introduced to the eNose a known mixture of all the palette odorants. Fig. 2 shows an example of such an experiment with a binary palette in which 162 different mixtures of the palette odorants were introduced to the eNose. In the overall, WSM2M predicts the concentrations rather accurately, although some deviations are observed for 2,3-butanedione in experiments 84–122. Most of the time the prediction errors are well below 5%, with a maximum of 8.9% for 2,3-butanedione and 10.9% for amyl formate.

The results of the entire experiment can be summarized by the mean value of the prediction error, in this case 0.71% (1.52%) for 2,3-butanedione, and 1.47% (2.19%) for amyl formate, with standard deviations in parentheses. The mean, however, is not a good measure of central tendency, since the distribution of prediction errors can be seen to have a strong positive skewness. Hence, the median would be a more stable measure of central tendency than the mean. In this particular example, the median values are 0.19% for 2,3-butanedione, and 0.59% for amyl formate. Hereinafter, we shall report median prediction errors only.

Median values of the prediction error for the entire set of experiments are shown in Fig. 3. The highest median values for binary, trinary and quaternary palettes are 1.68%, 6.81% and 8.11%, respectively. This suggests degradation in performance with palette size, but larger palettes should be tested in order to verify this hypothesis. The single quinary palette that we tested does not follow this trend, giving highest median value of 4.55%.

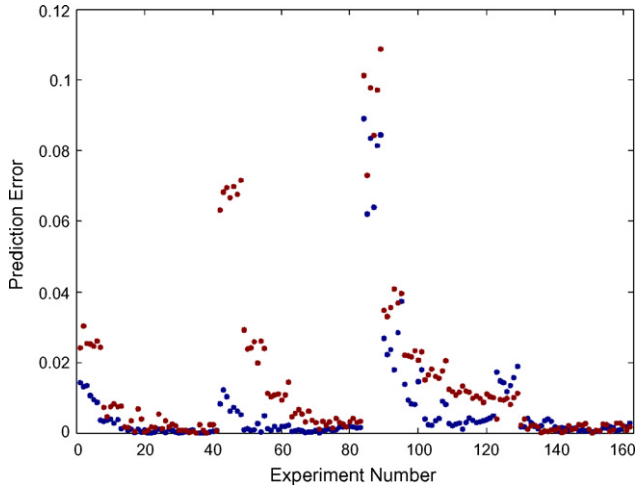

Fig. 2. Prediction error for a palette comprising 2,3-butanedione (blue) and amyl formate (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

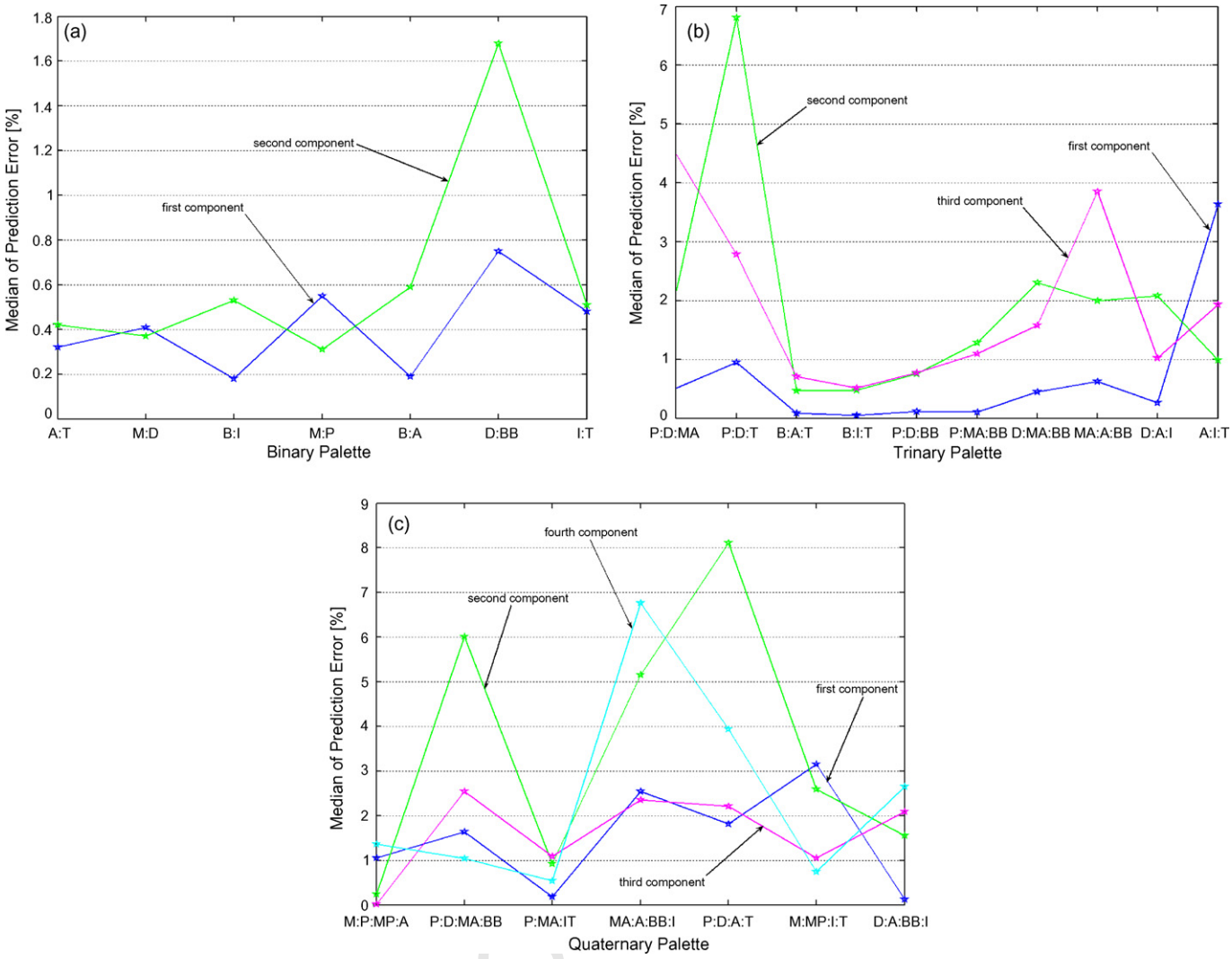

Fig. 3. Median values of prediction error, computed for each palette odorant. The WSM2M algorithm was applied to (a) 7 binary palettes, (b) 10 trinary palettes, (c) 7 quaternary palettes and one quinary palette (see text), and tested against known mixtures of the palette odorants.

However, these data should be interpreted with caution due to the single palette used. Recall our mentioning earlier that the reproducibility of our eNose is estimated to be in the range of 5–10%, and thus the performance of WSM2M in reconstructing mixtures of the palette odorants seems to be in agreement with this estimate.

5.3. Partial mixtures reconstruction

In a second set of experiments we used trinary, quaternary and quinary palettes, but used mixtures made of a part of the palette odorants. This allows one to estimate the amount of unnecessary ingredients WSM2M puts into its predicted mixture. The median values of the prediction error are shown in Fig. 4.

The results suggest that we should examine separately the prediction error of components present in the mixture (present-components) and that of components that are absent from the mixture (absent-components). Tables 2 and 3 show the different quality of prediction in these two groups. Application of the unpaired two-sample t-test ($P < 0.0001$), and the Wilcoxon rank sum test ($P = 0.0005$) shows that the distribution of median values of prediction error is significantly different between present-components and absent-components.

The WSM2M algorithm seems to work pretty well on present-components, with only two cases for which the median prediction error exceeded 10%. However, some "noise" is added to the predicted mixtures in the form of the exaggerated contribution of absent-components.

Table 2
Histogram of median values of prediction error for palette odorants that participate in the mixture that was actually introduced to the eNose

Palette	0–5%	5–10%	10–15%	15–20%	20+%
Trinary	15 (94%)	0	1 (6%)	0	0
Quaternary	29 (88%)	3 (9%)	0	1 (3%)	0
Quinary	7 (87.5%)	1 (12.5%)	0	0	0

Counts are shown in each entry, with the corresponding percentage in parentheses.

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

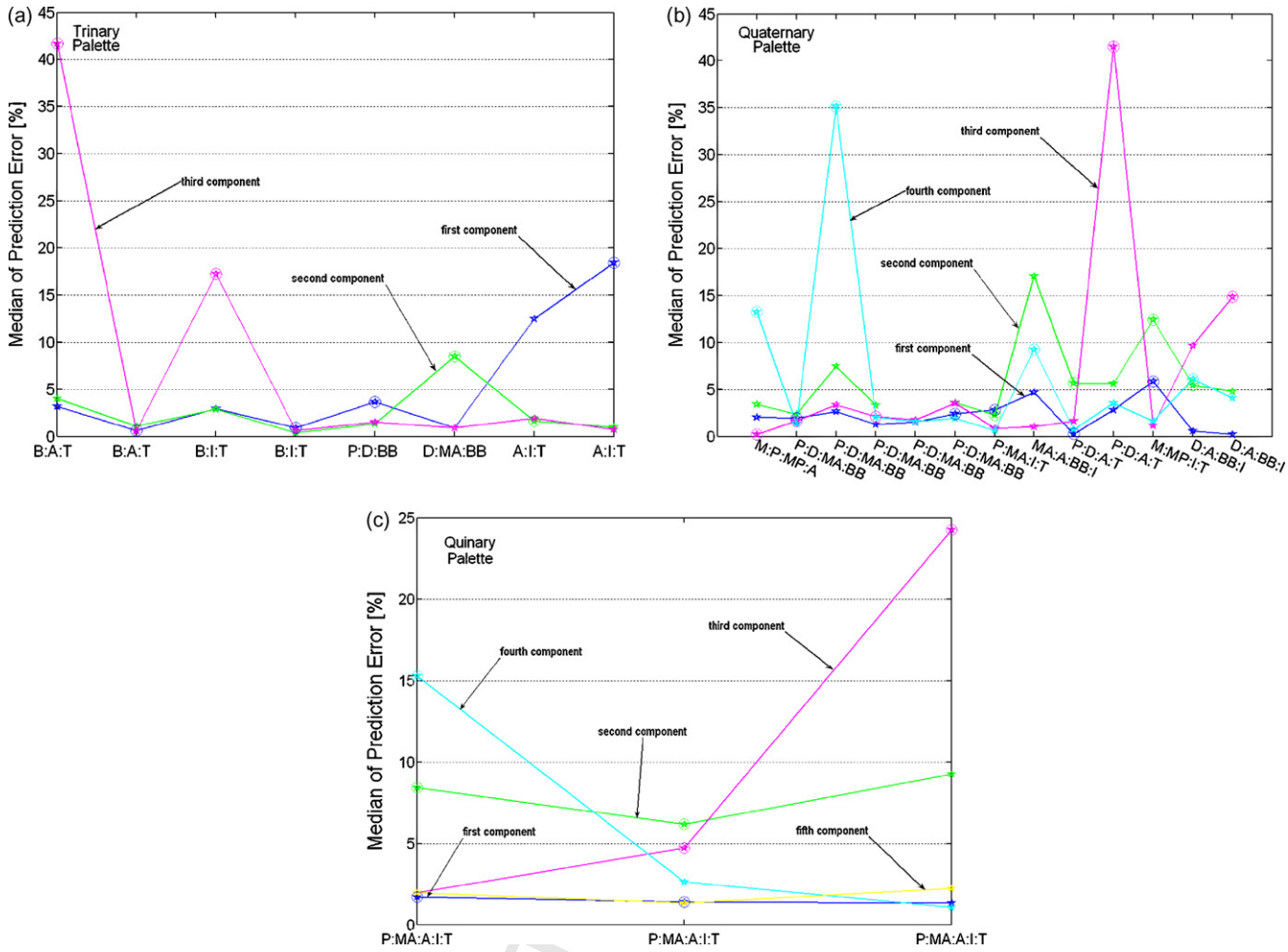

Fig. 4. Median values of prediction error, computed for each palette odorant. The WSM2M algorithm was applied to (a) 8 different combinations of partial trinary palettes, (b) 13 different combinations of partial quaternary palettes, and (c) 3 different combinations of partial quinary palettes. Those palette odorants that are missing from the tested mixtures are encircled.

Table 3
Histogram of median values of prediction error for palette odorants that are absent from the mixture that was actually introduced to the eNose

Palette	0–5%	5–10%	10–15%	15–20%	20+%
Trinary	4 (50%)	1 (12.5%)	0	2 (25%)	1 (12.5%)
Quaternary	8 (44%)	5 (28%)	3 (17%)	0	2 (11%)
Quinary	3 (43%)	2 (29%)	0	1 (14%)	1 (14%)

Counts are shown in each entry, with the corresponding percentage in parentheses.

6. Discussion

Ultimate validation of the WSM2M algorithm would involve doing the following for a series of arbitrary odorants. For an odorant eliciting response r from the eNose, use the algorithm to predict the palette mixture $\hat{v}$ for r. Introduce $\hat{v}$ to the eNose and compare the response to r. Unfortunately, for various technical/logistic reasons our laboratory work had to be terminated before we were able to do this. As a second best approach, we used the results we had for partial mixtures of the palette odorants to draw our conclusions.

Algorithmic reconstruction of mixtures of all palette odorants – also investigated by Nakamoto et al. [2,3] – seems to work quite well. As the palette size increases, the mean of the median prediction errors goes up too; see Supplementary Fig. S1. This is probably qualitatively true, but we cannot draw any quantitative conclusions, since the number of palette sizes we checked is small, and only a few experiments were carried out using the quinary palette. We anticipate that the curve depicted in the figure will eventually converge as the palette size grows (hopefully, to a sufficiently small prediction error).

For odor communication recall that the palette is chosen only once (or once per type of application), and we have full control over its contents. The average quantities quoted above can thus be taken as indicators only, since individual palettes can deviate significantly from them, for better or for worse. For example,

Please cite this article in press as: L. Carmel, D. Harel, Mix-to-mimic odor synthesis for electronic noses, Sens. Actuators B: Chem. (2007), doi:10.1016/j.snb.2007.03.022

Fig. 3 shows that while the binary palette D:BB is particularly "bad", palettes A:T and M:D are "good". Similarly, the trinary palettes B:A:T, B:I:T and P:D:BB look much better than other trinary palettes, having median prediction errors of less than 1% for all components. The same applies to quaternary mixtures M:P:MP:A and P:MA:I:T, which seem to exhibit particularly good reconstructions; all median prediction errors are well under 2%. Thus, specific palettes show better performance than average performance of same-size palettes, so these are the ones that should be used in an odor communication system.

There are probably numerous factors determining the performance level of a particular palette. For example, a strong (non-linear) interaction between palette's odorants is expected to decrease the performance. In any case, for WSM2M we will use those palettes that show, empirically, the lowest prediction errors. For the full algorithm, M2M, additional considerations might apply, such as different detection thresholds of the components, different levels of chemical stability, and the psychophysical impact of the odors on human sniffers (see also discussion in Ref. [15]).

Reconstructing mixtures for only some of the palette odorants is more difficult. We have demonstrated that the reconstructing process differs for present-components and absent-components (noise), being much worse for the latter. Still, the logic above applies here too, since susceptibility of the algorithm to noise is palette-dependent: The quaternary mixture P:D:MA:BB participated in 5 different experiments (totaling 14 present- and 6 absent-components), but only on one occasion showed a median prediction error greater than 10% (Fig. 4).

For each major application the design of the palette should be made only once, based on experiments like the ones shown here, but extended to larger palettes. For the palette chosen, another comprehensive set of experiments, like the ones described in Carmel et al. [15], should be carried out, to determine the mixing coefficients and response curves of the palette odorants.

This concludes our efforts for a fully functioning within-sniffer M2M algorithm. However, recall that it is but the first of three increasingly complex algorithms required for the full odor communication system of Harel et al. [1].

We would like to briefly describe our work on the other two algorithms. The BSM2M algorithm adds a further complication, requiring mapping response vectors in one eNose to their corresponding response vectors in a second one. As the two might feature sensors of completely different nature, this is far from trivial. We have shown the feasibility of such a mapping using one eNose based on quartz crystal microbalance sensors and another on conducting polymers. We found that while the global topologies of the feature spaces are very different, local topologies are remarkably preserved [20]. This enabled us to design a local mapping technique, tessellation-based linear interpolation, which yielded an accurate mapping. The full M2M scheme requires a more difficult mapping, from response vectors to perceptions, whose existence has not yet been proven, although evidence shows that such a mapping does exist [1].

Uncited reference

[21].

Acknowledgement

Many thanks to Doron Lancet who cooperated with us in the preliminary stages of the work, and with whom we have had numerous enriching discussions over the years.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.snb.2007.03.022.

References

[1] D. Harel, L. Carmel, D. Lancet, Comput. Biol. Chem. 27 (2003) 121–133.
[2] T. Nakamoto, Y. Nakahira, H. Hiramatsu, T. Moriizumi, Sens. Actuators B: Chem. 76 (2001) 465–469.
[3] T. Yamanaka, R. Matsumoto, T. Nakamoto, Sens. Actuators B: Chem. 89 (2003) 112–119.
[4] K. Persaud, G. Dodd, Nature 299 (1982) 352–355.
[5] L. Buck, R. Axel, Cell 65 (1991) 175–187.
[6] J.J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 12506–12511.
[7] L. Carmel, N. Sever, D. Lancet, D. Harel, Sens. Actuators B: Chem. 93 (2003) 77–83.
[8] J.W. Gardner, H.W. Shin, E.L. Hines, Sens. Actuators B: Chem. 70 (2000) 19–24.
[9] M.E. Monge, D. Bulone, D. Giacomazza, D.L. Bernik, R.M. Negri, Sens. Actuators B: Chem. 101 (2004) 28–38.
[10] R. Dutta, E.L. Hines, J.W. Gardner, K.R. Kashwan, M. Bhuyan, Sens. Actuators B: Chem. 94 (2003) 228–237.
[11] M. Penza, G. Cassano, Food Chem. 86 (2004) 283–296.
[12] B.J. Hwang, J.Y. Yang, C.W. Lin, Sens. Actuators B: Chem. 75 (2001) 67–75.
[13] C. Di Natale, A. Macagnano, S. Nardis, R. Paolesse, C. Falconi, E. Proietti, P. Siciliano, R. Rella, A. Taurino, A. Amico, Sens. Actuators B: Chem. 78 (2001) 303–309.
[14] M. Pardo, G. Faglia, G. Sberveglieri, M. Corte, F. Masulli, M. Riani, Sens. Actuators B: Chem. 65 (2000) 267–269.
[15] L. Carmel, N. Sever, D. Harel, Sens. Actuators B: Chem. 106 (2005) 128–135.
[16] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood, 1974.
[17] J. Mitrovics, H. Ulmer, U. Weimar, W. Gopel, Acc. Chem. Res. 31 (1998) 307–315.
[18] J.W. Gardner, P.N. Bartlett, Electronic Noses: Principles and Applications, Oxford University Press, Oxford, 1999.
[19] H.T. Nagle, R. Gutierrez-Osuna, S.S. Schiffman, IEEE Spectrum 35 (1998) 22–31.
[20] O. Shaham, L. Carmel, D. Harel, Sens. Actuators B: Chem. 106 (2005) 76–82.
[21] R.C. Araneda, A.D. Kini, S. Firestein, Nat. Neurosci. 3 (2000) 1248–1255.