Sensors and Actuators B 231 (2016) 666–674

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

journal homepage: www.elsevier.com/locate/snb

Temporal responses of chemically diverse sensor arrays for machine olfaction using artificial intelligence

Shaun K. Ryman [a], Neil D.B. Bruce [b,**], Michael S. Freund [a,*]

[a] Department of Chemistry, University of Manitoba Winnipeg, MB, R3T 2N2, Canada
[b] Department of Computer Science, University of Manitoba, Winnipeg, R3T 2N2, Canada

ARTICLE INFO

Article history:
Received 11 July 2015
Received in revised form 18 February 2016
Accepted 15 March 2016
Available online 17 March 2016

Keywords:
Olfaction
Sensor array
Pattern recognition
Machine learning
Machine sensing

ABSTRACT

Machine sensing has progressed through innovative breakthroughs. Perhaps the most significant advancements have been in the field of machine vision through the advent of the charge-coupled-device (CCD) chip and the development of artificial neural networks. This is evident in the range of sophisticated behaviors that have been replicated including facial recognition and object tracking. Due to parallel processing of large amounts of temporally evolving data, these systems can rapidly learn, adapt and predict in changing environments. While there have been significant advances in chemically diverse sensor arrays that mimic the mammalian olfactory system under well-defined conditions, relatively little work has been done to understand the temporal response of these systems and to develop neural architectures that will enable machines to become chemically aware of their surroundings. In this work, the temporal response behavior for sensors is investigated, demonstrating that both the response evolution and steady-state provide useful information about the chemical environment under dynamic conditions and that neural networks are ideally suited to combine this information in unsupervised learning to perceive odors in a dynamic environment.

1. Introduction

Adaptability and learning is essential for the survival and evolution of organisms in a changing environment. Artificial neural networks (ANNs) can replicate the plasticity and versatility of natural neurons as well as, their ability to perform complex tasks and to retain information [1–4].

While significant advancements within the field of ANNs have been made within the past few decades, their true potential has yet to be fully realized [1]. Many advancements have been inspired by natural systems both in how they function and the tasks they are trained for [5]. In the realm of machine vision, significant progress has been made for a variety of recognition problems bolstered by the role of GPUs in accelerating learning, and by the proliferation of large scale data sets. Moreover, there has been recent success in integrating visual information with language to allow for rich English descriptions of image content [6]. In contrast, a relative paucity of efforts have addressed the problem of machine olfaction from a vantage point that leverages opportunities availed by deep learning. Some of the challenges that belie further development in this domain include developing a better understanding of sensor behavior for olfaction, and identifying some of the problems that are unique to this landscape. As such, this marks a very rich area for future research that is relatively unexplored. On the horizon, progress in machine olfaction has the potential to provide an information rich data source for reasoning, and applications that parallels the success that has been observed for machine vision. With this, opportunities for multi-modal sensing also stand to significantly raise the ceiling on sensory capabilities in intelligent systems.

The first ANN was designed using mathematical models to mimic the "all or none" response from neural cells [7]. Investigation and development of this idea has lead to expansion into new architectures and propagation schemes [1]. A particularly fruitful area has been in the integration of ANNs with sensor arrays. For example, there has already been heavy implementation with CCD chips

Abbreviations: CCD, charge-coupled-device; IF, intensity Factor; PCA, principal component analysis; ANN, artificial neural network; PVP, poly(*n*-vinyl pyrrolidone); PSA, poly(styrene-co-allyl alcohol); PVB, poly(vinyl butyl); PSMA, poly(styrene-co-maleic anhydride); PS, poly(sulfone); PMMA, poly(methyl methacrylate); PMVEMA, poly(methyl vinyl ether-co-maleic anhydride); PEO, poly(ehtylene oxide).

* Corresponding author at: Department of Chemistry, Florida Institute of Technology, 150 West University Boulevard, Melbourne, FL 32901-6975, USA.
** Corresponding author at: E2-408, Department of Computer Science, University of Manitoba, Winnipeg R3T 2N2, Canada.
E-mail addresses: Bruce@cs.umanitoba.ca (N.D.B. Bruce), msfreund@fit.edu (M.S. Freund).

http://dx.doi.org/10.1016/j.snb.2016.03.059
0925-4005/

for face and object recognition. While ANNs have been explored in the context of olfaction [8], the chemical sensing equivalent of the CCD chip has yet to emerge. The value of the representation of odors expressed by a sensor array resides in the richness of odor specific variation expressed by the sensors, by the sophistication of reasoning, and pattern recognition applied to the raw sensor behavior. The coupling of a rich sensor array with modern developments in neural networks opens up a number of exciting opportunities such as real-time analysis of multiple gaseous analytes in dynamic environments.

Significant advances have been made in the creation of chemically diverse sensor arrays, which may eventually serve this purpose [9]. For example an array of chemically different polymer resistance sensors can generate response patterns in a manner analogous to the natural system where identification is achieved from a response pattern generated by an array of receptors within the olfactory epithelium.

Carbon black-organic polymer resistance sensors are particularly attractive due to their ease of manufacturing and low cost as well as their fast, reproducible responses [10]. It has been shown that when equilibrated with vapor the resistance grows within the sensor due to the absorption of the gas within the carbon black-organic polymer matrix increasing the distance between the conductive carbon black particles [11]. The degree of absorption is a function of the nature of the polymer in the composite and as long as the time constant of the sensor, associated with diffusion of the vapor in the sensing layer, is sufficiently fast, the sensor resistance will track the concentration of the vapor as a function of time. To date, these types of sensor arrays have been limited to highly controlled flow systems and have not transitioned into real and dynamic environments, nor have they been integrated with sophisticated ANNs that will enable their progression into mass produced technology or amalgamation into the internet of things.

In order to make significant advances in machine learning around chemical sensing, there are many complex features that have to be replicated. For example, the ability to discriminate odors in a complex and changing environment. Until now, the focus has been on well-defined analyte delivery systems to produce equilibrated responses with relatively simplistic multivariate analysis. The separation of analyte specific response patterns and intensity is essential for the dual task of identifying and quantifying in more complex environments. The response of a sensor array to a given analyte depends on the partitioning of the analyte vapor into each sensor, which is a function of the nature of the organic polymer and the analyte as well as mass transport [12]. The intensity of the response pattern is then related to the concentration or vapor pressure of the analyte.

Dimensionality reduction using ANNs may be used to decouple intensity from the characteristic response patterns. The advantage of ANNs over more standard methods involving linear transformations such as principal component analysis (PCA), is that data points may be characterized subject to their proximity to a (possibly curved) manifold that resides in a subspace of the signal measurement data. Moreover, there is flexibility in constraining the properties of a compressed representation; often activation among neurons in a compressed representation is penalized at the learning stage to promote sparsity in how observed signals are represented in the network.

Autoencoders are perhaps the most standard ANN used for encoding signals [13]. In an autoencoder, an initial layer of neurons that matches the dimensionality of signal measurements (number of sensors in this case) is passed on to subsequent layers with fewer and fewer units providing a more compressed representation of the signal. Weights in the network are adjusted such that the error in reproducing the input data from the most compressed layer is minimized.

Table 1
List of polymers used, chemical source and their molecular weights.

Polymer	Acronym	MW
Poly(vinyl butyral), from Polysciences	PVB	70,000
Poly(styrene-co-allyl alcohol), from Sigma Aldrich	PSA	2200
Poly(*n*-vinyl pyrrolidone), from Accros	PVP	58,000
Polysulfone, from Sigma Aldrich	PS	22,000
Poly(methyl methacrylate), Sigma Aldrich	PMMA	120,000
Poly(methyl vinyl ether co maleic anhydride), Sigma Aldrich	PMVEMA	216,000
Poly(styrene-co-maleic anhydride), Sigma Aldrich	PSMA	224,000
Poly(ethylene oxide), Sigma Aldrich	PEO	100,000

Sparse autoencoders extend this idea by also imposing a penalty that depends on the activation of units at each layer in representing input patterns. This will tend to result in a relatively small number of neurons being active at a time for different patterns that are observed. There is also considerable evidence that neural information processing in biological systems is structured in this manner [14], and sparse autoencoders may produce neurons with properties reminiscent of those appearing in the human brain for early visual [15] or auditory and somatosensory [16] representation. A further extension of autoencoders is the sparse filtering method [17]. Sparse filtering differs from a standard sparse autoencoder in that each observation (set of sensor measurements) is normalized to have a unit standard deviation, and normalization of the same form is also applied to all measurements from a single observation over all samples (a single sensor over time) exhibiting a similar effect as normalization with the intensity factor described above. This type of normalization has been shown to produce representations that achieve better performance for pattern classification. In addition, this normalization is within a mathematical form that matches the form of apparent canonical normalization observed in sensory systems in humans and other species [18].

In this work, the focus is on exploring the temporal behavior of chemical sensor arrays and the relationship between response patterns and intensity. This understanding is then used to facilitate the design of ANNs that are suited for processing and learning from these sorts of sensor arrays under dynamic conditions. The setup associated with this line of experimentation allows for acquisition of new data sets, variation in number and type of sensors, and observations of intrinsic sources of variation and noise associated with the sensor array. In moving forward, this will provide the possibility of developing richer data sets that allow for more complex neural models to be exercised for reasoning. It is also expected, that understanding of high-level behavior and function of sensor array behavior will provide insights that may be related to alternative sensor types or configurations. At this stage, observation of some of the rudiments of sensor behavior are sufficient to dictate suitable architectural configurations among neural networks to accommodate for some of the challenges specific to olfactory sensing.

2. Materials and methods

Sensors were made using a ratio of 40 mg Carbon black (Black Pearls 2000, generously donated by Cabot Corp.) to 160 mg of organic polymer in 20 ml of solvent. The 8 Organic polymers used include poly(vinyl butyral), poly(styrene-co-allyl alcohol), poly(*n*-vinyl pyrrolidone), polysulfone, poly(methyl methacrylate), poly(methyl vinyl ether-co-maleic anhydride), poly(styrene-co-maleic anhydride) and poly(ethylene oxide) (see Table 1). The solution was then sonicated for 15 min and airbrushed onto an interdigitated electrode (IDA) using an aluminum mask. Resistance

was measured using an ohmmeter while airbrushing until the resistance was below 100 kΩ for each sensor and a black film was observed. There are 7 sensor elements within one IDA chip. Each chip therefore consists of 7 replicate sensor elements. IDA's were then allowed to dry for 24 h before analysis. The main description of the IDA as well as instrumental setup of the olfactometer are described elsewhere [19].

Depending on the experiment the vapor generation procedure was altered, there are 4 main procedures used within this work. This includes repeated exposures, random exposures, concentration ramping exposures and gradual ramping exposures. Repeated exposure and random exposures the sensor array is exposed to a nitrogen background for one hour at 400 standard cubic centimeters per minute (sccm). Followed by a 20 min exposure of the analyte at 5% P/P° then 20 min of the background 5 times with 60 min in between vapors (this is omitted for random exposures). Sensors were broken by a 20-minute on/off exposure to each analyte before each experiment. This was a single exposure to each vapor followed by a 60 min exposure to the background vapor.

The main experimental setup involved an olfactometer. Electronic measurements were performed using a 34980A Multifunction Switch/Measure Unit from Agilent with a 3932A 80 channel reed multiplexer with two point resistance measurements taken either every 15 s or 2 min with a Gain (M) of 1, Offset (B) of 0, Res of 5.5 and command execution times under 6 millisec for GPIB. Gas flow was controlled using a mass flow control system made by MKS technology (model #2179A) each independently controlled by the computer and the Flowcon (built by Plasmoinique Inc., St Hyacinthe, PQ). The mass flow controllers have three-way valves and Teflon solenoid shut-off valves. They are controlled through the PC and can independently set each vapor to the desired vapor pressure from 0.2 to 2000 sccm [20].

Nitrogen was the carrier gas for the 5 analytes measured, which include toluene, dichloromethane (DCM), methanol, acetonitrile (AN), and water for a total flow of 400 sccm and 420 sccm for the gradual ramping experiments. Sensors were initially exposed to 60 min of N_2 gas to reach a stable baseline, then subsequent exposures of 20 min on/off to each vapor to a total of 5 repetitions per analyte.

Processing of the experimental data is done using PCA and then transitioning to ANN for pattern analysis. ANN will use both unsupervised and supervised methods with a sparse autoencoder as backpropagation and feedforward neural networks respectively. Low-dimensionality coding allows the discrimination of different analytes to be observed directly, quantitatively and create relatable visual comparisons to PCA.

Sensor inputs project onto a layer of neurons by way of a linear transformation as follows:

$f_j^{(i)} = w_j^T x^{(i)}$ with $x^{(i)}$ corresponding to input pattern i (a vector of sensor response values), and w_j corresponding to the weights associated with index j.

Outputs derived from linear filtering are subject to an L2-norm both across the cell population, and across the sample population:

$$\bar{f}_j = f_j / \|f_j\|_2$$

$$\hat{f}^{(i)} = f^{(i)} / \|f^{(i)}\|_2$$

This normalization is responsible for dispersal across individual neurons following the linear transformation, and lifetime sparsity of cells so that all units represent meaningful information across the stream of inputs that are processed.

Finally, optimization is done according to the constrained optimization of the expression:

$$minimize \sum_{i=1}^{N} \|\hat{f}\|_1$$

This is solved by way of standard L-BFGS optimization, and the L_1 norm of the array of neurons coupled with L_2 normalization described above yields a meaningful separation of vapors, and relative invariance to concentration.

In the experiments presented in the paper, the input array is reduced by Sparse Filtering to a set of 12 neurons. For visualization, this is subsequently reduced to 3-dimensions with the 3 principal dimensions depicted in the scatter plot.

Results appearing in Fig. S9 are produced by a feedforward neural network. The input layer consists of 63 sensor response values for 5 sequential time samples (315 dimensions). Two hidden layers of 18 and 6 neurons respectively follow, with one neuron at the output layer per category of vapor. All layers are fully connected to adjacent layers. Error is based on mean-squared-error, and Levenberg–Marquardt optimization is used in training the network. For visualization, activation at the output layer on the test data is reduced to 3 dimensions via PCA with the scatter plot showing values corresponding to the first 3 principal components.

The Davies-Bouldin index (DBI) is used to analyze how vapor samples are clustered in the original sensor coordinate space relative to the encoding defined by sparse filtering. This is not a traditional clustering problem per se, as cluster labels are not being assigned to data points but rather labels are defined by the vapor sample. Nevertheless, the notion of how well clustered these points are is analogous to the problem faced in evaluating results of clustering. Note that normaliztion based on cross sample and cross sensor L^2-norm applied in sparse filtering is also applied to the raw sensor data to place these values on an equivalent numeric scale. The DBI is based on the ratio of within-cluster scatter to inter-cluster separation, or more specifically:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{j \neq i} \left\{ D_{i,j} \right\}$$

Where $D_{i,j}$ is the within-to-between cluster distance ration for the i^{th} and j^{th} clusters.

$$D_{i,j} = \frac{\left(\bar{d}_i + \bar{d}_j\right)}{d_{i,j}}$$

Where $\bar{d}_i$ is the average distance between each point in the ith cluster and the centroid of the ith cluster, $\bar{d}_j$ is the average distance between each point in the jth cluster and the centroid of the jth cluster and $d_{i,j}$ is the Euclidean distance between the centroids of the ith and jth clusters.

The maximum value of $D_{i,j}$ captured by DBI corresponds to the worst-case within-to-between cluster ratio for cluster i. The smallest Davies-Bouldin index value corresponds to the best clustering solution.

3. Results and discussions

3.1. Sensor behavior

3.1.1. Well-defined conditions

The design and implementation of a robust method for real-time quantitative vapor recognition requires a detailed understanding of sensor temporal performance. In particular understanding the range of response time constants and how they change as a function

of sensing polymer and analyte as well as the concentration dependence of response is essential for the design and implementation of ANNs for processing these types of data streams. Quantification is achieved by removing the intensity from the data set to leave a concentration independent component for classification.

Ideal conditions for sensor analysis is defined as standard temperatures of 25 °C using controlled exposures of analyte vapors, exposing until equilibrated response has been reached. Under well-defined conditions each sensor was tested using standard methods as described in previous work [9]. The sensor array was equilibrated for each exposure to 4 vapors of toluene, dichloromethane, methanol, and acetonitrile. The vapor analytes were chosen for the scale of solvation parameters consisting of excess molar refraction, solute polarizability-dipolarizability, hydrogen bonding acidity, hydrogen bonding basicity, and solute-gas hexadecane partition coefficient as seen in Table S3 and determined by Abraham et al. [21,22]. Water was not used in this study since sensitivity was low due to the hydrophobic nature of the carbon black in these particular sensors.

Principal component analysis was used as a simple method for pattern recognition and visualization for multidimensional data results from a sensor array of 8 unique sensors with several duplicates of each (55 sensors total due to one unresponsive sensor). Under well-defined conditions, with a set flow, constant vapor pressure and using on/off exposures, the sensor array exhibits reproducible response patterns [23]. A 20 min exposure permitted all the sensors to reach equilibrium producing a maximum ΔR/R response to a particular analyte and returning to the baseline after 20 min exposure to carrier gas in the absence of analyte.

Increases in analyte concentration consequently results in a corresponding increase in ΔR/R response for the sensor array (see Fig. S1). With increasing concentration the response pattern also increases, due to an increase of partitioning into the sensor and a corresponding increase in resistance (see Fig. 1A) [10]. Within principal component space an increase in concentration results in the increase in principal component values along trajectories unique to each analyte. Linear principal component behavior is consistent with a linear response for each individual sensor and leads to the expectation that the summation of the responses will also be linear and can therefore be used as an intensity factor. This summation or intensity factor can in turn be used as a single concentration dependent parameter (see Fig. S2), which can be used to normalize each response to create a concentration independent response pattern (see Fig. S3).

$$IntensityFactor\,(t) = \sum \Delta R/R\ Vapor\ (t)$$

This method of normalizing to separate the data into its concentration independent component, where ΔR/R is the change in response of a specific sensor with respect to its baseline and $\sum$ΔR/R is the sum of the responses from each sensor within the array for the analyte at any point in time.

At low concentrations the normalized data exhibits significant fluctuations due to the decreased signal-to-noise of both sensor responses and the intensity factor, however a threshold can be identified above which the patterns become stable which for methanol occurs above 3% P/P° in PC1 (see Fig. S3). Normalization using the intensity factor produces a concentration independent response pattern (see Fig. S4). This is also observed in principal component space where at lower concentrations there is large scatter for each vapor and depending on the sensitivity of each sensor to the sensor array the points cluster after a given concentration (see Fig. 1B and Fig. S5). This clustering indicates recognition of a given pattern and classification of the analyte vapor is possible independent of concentration. The thresholds for each vapor were determined to be 0.8, 11.8, 4.0, 5.6 ppth at 25 °C and 1 atm for

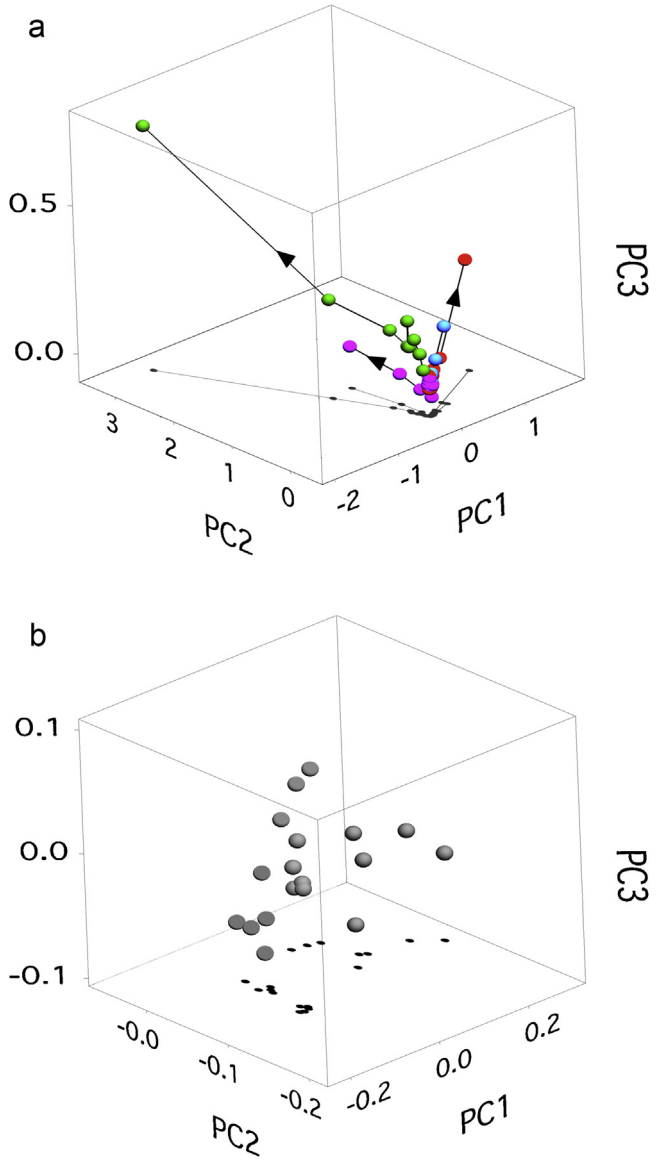

Fig. 1. (a) PC1, PC2, and PC3 of the ΔR/R response of the sensors array with increasing concentration of toluene(blue), dichloromethane(red), methanol(green) and acetonitrile(purple). Increasing concentration is shown with the black arrows originating from (0,0,0), concentrations are 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 5.0, 10.0 and 20.0% P/P°. (b) PC1–PC3 of the Normalized ΔR/R response of the sensors array with increasing concentration. Points below 0.5, 1.0 and 1.5% P/P° have been omitted from the image because they are randomly scattered within the principal component space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

toluene, dichloromethane, methanol and acetonitrile respectively. The detection limits were found using the normalized response pattern at the varying concentrations. When the sensor array is above a threshold concentration, aggregates form within the clustered data in principal component space it is said to be responding well. For all sensors a reasonable limit was found to be at 5% P/P°, this corresponds to 1.7, 19.7, 7.9 and 5.6 ppth for toluene, dichloromethane, methanol and acetonitrile respectively.

3.1.2. Dynamic system

In real-world conditions there will be changing vapor concentrations that will introduce new complexity to the analyses depending on the rate of change in the environment and the time constant of the sensors. A dynamic system was simulated by gradually changing the concentration of a vapor within a nitro-

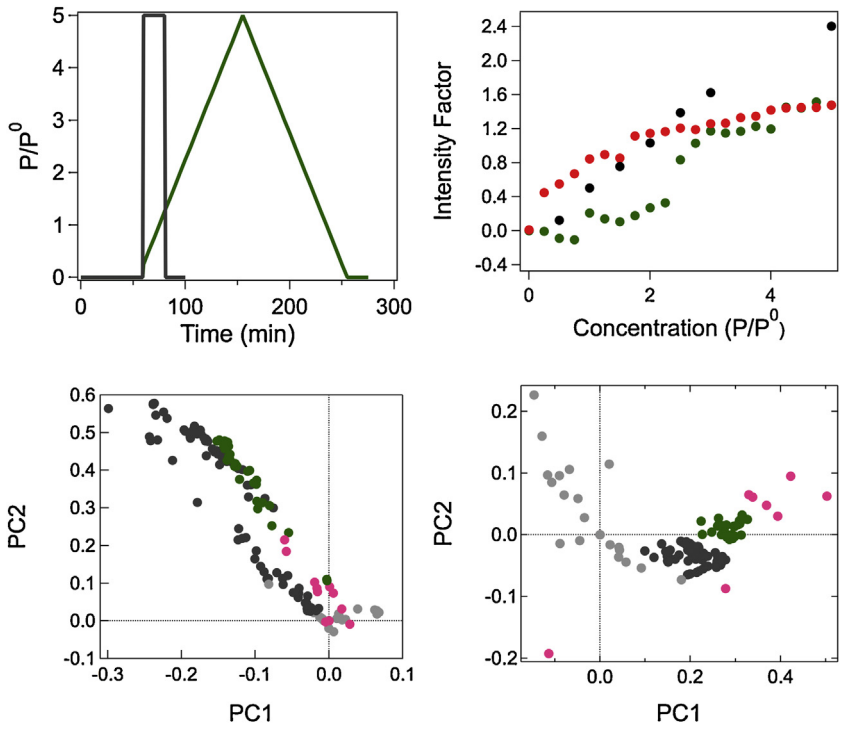

Fig. 2. Comparison of stepped versus gradual change in concentration for methanol. (a) Partial pressure applied over time to the sensor array. The concentration was gradually raised to a maximum of 5% P/P° at a rate of 0.25% P/P° every 5 min (b) Intensity factor changes with concentration. Rise (green) $R^2 = 0.921$, fall (red) $R^2 = 0.847$. (c) PC1 and PC2 response of $\Delta R/R$. d) Response of PC1 and PC2 for normalized data using the Intensity factor. Noisy, randomly scattered data points below the threshold intensity are marked in pink. Significant variance ensues below 2.5% P/P° (3.95 ppth) at 25C° and 1 atm. Points at (−0.049, 1.20) and (6.56, 2.73) are randomly scattered within principal component space and have been omitted from the image for a better visual of the cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

gen gas source. The effects of exposure mimics what would be expected when the sensor array slowly approaches or moves away from an odor source or the gradual occurrence or dissipation of an odor source (see Fig. 2A). Analysis of the transient exposure demonstrates the applicability of the normalization approach when comparing methanol in the principal component space. The response extends outwards linearly in space as the concentration increases, then retreats back to the origin as the partial pressure is reduced (see Fig. 2B and C). Once the intensity surpasses the threshold for identification the representation within principal component space switches from a random scatter into a concentrated cluster of points, simplifying the classification of analytes (see Fig. 2D).

Changing concentration adds to the complexity of identifying an analyte since these changes alter the response location within principal component space. As shown above, under static conditions the pattern remains the same and can be normalized with an intensity factor. Under dynamic conditions where the rate of equilibration (or time constant) of each sensor will be dependent on the nature of both the vapor and the polymer, the temporal behavior will be more complex. For example, polyethylene oxide (PEO) containing sensors reach equilibrium for methanol within 1–2 min whereas poly(methyl vinyl ether-alt maleic anhydride) (PMVEMA) equilibrates more slowly in 5–6 min (see Fig. S6).

The complexity in the temporal response in the 4 vapors tested exhibit fairly intricate but different response loops in principal component space (see Fig. 3A). Rate of vapors absorption into the sensors, and the corresponding changing resistance, can be described using time constants [24]. The modelled resistance response can be used to determine the time constant from the rate of sensor saturation. Comparisons of the sensor time constant responses range from fast to slow where 1 time constant is equivalent to 63.2% saturation (see Table S1). The faster responding sensors (PSA, PEO, PSMA, PS, PVB) reach their saturated $\Delta R/R$ response more quickly than the slower sensors (PVP, PMMA, PMVEMA) to a given analyte such as Methanol. Normalized sensor array response patterns near the baseline of a peak are randomly scattered (see Fig. S7), but above a given threshold cluster together within principal component space (see Fig. 3B). As patterns evolve they maintain their unique character despite differing time constants. A similar event is observed within Fig. 1B, the difference being the way the calculation was carried out, the results were calculated using the maximum $\Delta R/R$ response to maintain the orientation of the data points within principal component space.

Variations in time constants occur both as a function of vapor for a single sensor type and as a function of sensor type for a single vapor (see Table S1). This is not surprising since each time constant is dependent on the interaction between the polymer and the vapor analyte as described by Abraham [21].

3.2. Artificial neural networks

3.2.1. Architecture and response

Unsupervised methods were chosen to provide a more autonomous process of odor classification which naturally lead to ANN compression of signals and sparse filtering which produces a representation with desirable properties that are not produced by PCA, autoencoders, or sparse autoencoders in their traditional form. In particular, embedding observed measurements in a 3 dimensional space results in clear separation of measurements corresponding to different vapors, and also relative invariance to concentration which is a hallmark of neural computation associated with olfaction in primates [25]. Normalization provides this invariance, and also diminishes sensitivity to sensor measurements when all of the sensors are relatively non-responsive. It is also worth noting that the normalization achieved is similar to that produced in

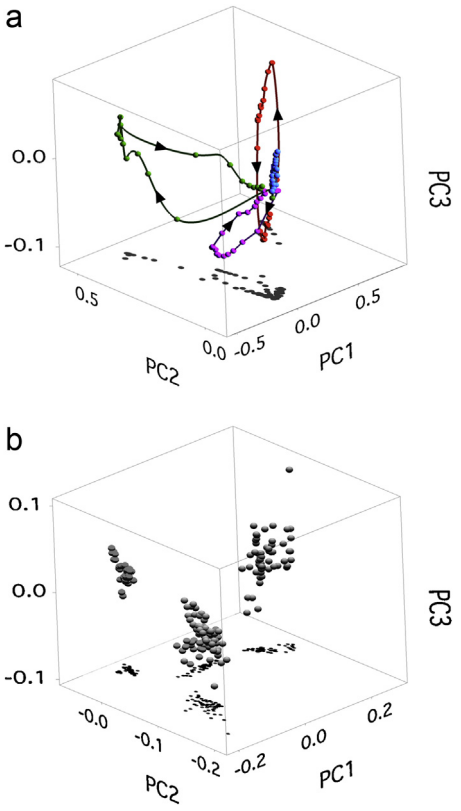

Fig. 3. (a) 3D plot of the averaged temporal response from the sensor array to the 4 vapors toluene(blue), dichloromethane(red), methanol(green) and acetonitrile(purple). Each point indicates 2 min up to the maximum $\Delta R/R$. (b) Normalized temporal response pattern from Fig. 3A using the Intensity Factor. Data points with an intensity factor close to zero have been omitted from the image because they are randomly scattered within principal component space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

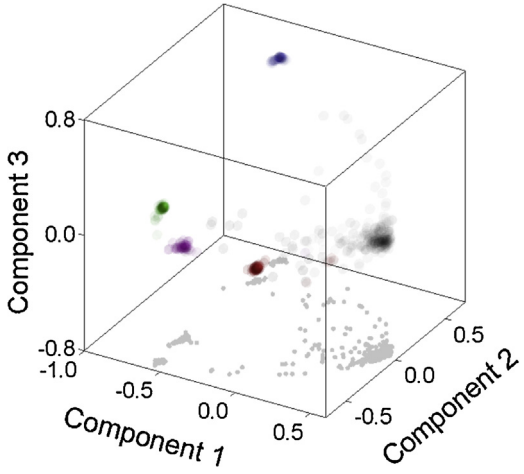

Fig. 4. Sensor array was exposed to 5.0% P/P° in 20 min on/off of toluene(blue), dichloromethane(red), methanol(green) acetonitrile(purple) and no analyte (grey) using ANN and a sparse autoencoder. A scatter plot is shown wherein the opacity of points plotted is a function of the sum of responses elicited by the sensor. This has the impact of making stronger measurements more visible, and also provides a stronger sense of point density in the projected space. Note that such scaling (or thresholding) may also be used as a part of the detection process in the form of half-wave rectification or a soft step function (e.g. sigmoidal). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

considering $\Delta R/R$ values, however in this case this results from the raw data, and basic principles tied to neural information processing that also have strong correlations with computation in the human brain. These characteristics provide a strong foundation for more involved processing using ANNs, as the number of sensors and vapors is scaled up, or in shifting to a supervised representation.

Given the separation appearing in Fig. 4, it is evident that a simple classifier may be highly successful for the data considered thus far. However, the problem of separating a much larger number of classes requires careful determination of class boundaries in the face of varying temporal dynamics (time constants) of sensors, time constants that vary as a function of vapor, hysteresis of the system on a shorter time scale, and changes to sensor properties on a longer time-scale. For example, experiments with concentration ramping yield less separable clusters when odors are encoded by sparse filtering, and show greater concentration related drift (Fig. S8). There is nevertheless some invariance to concentration given the clustering of samples in the encoded space which speaks to the value of normalization strategies in achieving desired invariance properties within the neural representation.

In considering the DBI corresponding to the unsupervised encoding of sensor responses, an error bound on these indices based on bootstrapping is also calculated. For all sensor response samples and their corresponding sparse filtering encoding, sampling with replacement is carried out respectively based on 10,000 samples to provide an estimate of the variance associated with the DBI for each case.

This quantifies the inter-cluster scatter relative to inter-cluster separation. Specifically each cluster has a different amount of inter-cluster scatter, and a different distance to the centroid of each other cluster. Values in Table 2 therefore correspond to the average worst case intra-cluster scatter to inter-cluster distance across all clusters as described in Section 2. Table 2 demonstrates the improvement in grouping of encoded vapor responses based on sparse filtering for the two different methods of exposure.

Tests with supervised training reveal the capacity for classification of vapors based on raw sensor array data (see Fig. S9). In the supervised case, a neural network with 315 units at the input layer, and 5 neurons at the output layer is used to classify vapors. The 315 units consist of 63 sensor responses for a sliding temporal window consisting of 5 samples. Two hidden layers were used between the input and output layers consisting of 18 and 6 neurons respectively. The network was trained on half of the sensor data, with the other half used to produce the plot appearing in the figure. In transitioning to solving this problem with a much larger pool of data, target vapors, and variable conditions, we anticipate that convolutional neural networks will be especially effective. This expectation is based on the ability to automatically determine appropriate temporal frequency bands for recognition to gain additional robustness to sensor drift and complex system dynamics.

Classic backpropagation for classification has been shown to benefit from initial seeding of weights derived using a sparse autoencoder. It is therefore the case that the desirable properties of sparse filtering for olfactory encoding are likely to extend to alternative neural network solutions even if only used for initialization in learning. In recent years, neural networks have provided very capable solutions for pattern classification (notably in machine vision), in employing network architectures with feedback and

Table 2
Calculated DBI values for unsupervised ANN clustering.

Data Analyzed	Repetitive Exposures	Concentration Ramping Exposures
Sensor Array	1.72 ± 0.13	22.14 ± 1.34
Sparse Filtered Encoding	0.60 ± 0.07	2.03 ± 0.10

memory. For example, Long Short Term Memory (LSTM) networks have demonstrated performance that exceeds human capabilities for classifying traffic signs. The structure of these networks effectively allows for storage and forgetting of states, and gating of signals on the basis of activation of neurons in the network. This provides a solution to addressing some of the challenges discussed in olfactory signal classification. Performance of LSTM and related architectures scales well with the number of categories [26], and systematic patterns of hysteresis may be accounted for by neurons that implicitly capture different network states. Moreover, robustness may be increased by learning of temporal dependencies both for vapors detected, and intermediate neurons that contribute to these detections. This is also true of the differing and vapor dependent temporal dynamics expressed across sensors.

4. Discussion

Our findings show the capability and behavior of sensor arrays within dynamic environments as well as to setup the beginning framework for advancements similar to those made for CCD chips, but to classify a selection of odors utilizing unsupervised techniques. The added complexity of changing vapor concentration and non-equilibrated responses provide much more information and realistic conditions but require more complex methods of processing, which we have demonstrated using ANNs.

The intensity factor minimizes drifting with concentration and confines the vapors to a given location within principal component space, effectively minimizing the concentration component of each analyte (see Fig. 2D). Once normalized the vapors at low concentration will appear as a random scatter (see Fig. S5), this is because a small $\Delta R/R$ is divided by a small intensity factor, which equals some larger number. Therefore a threshold for each vapor is proposed to remove the noise from the analysis, by paying attention to the vapors only above a given preset value (see Fig. 1B). With a threshold set, the trained sensor array is capable of monitoring for analytes with time (see Fig. 3A), selecting the vapor responses that are above a given threshold (see Fig. 3B). Without the threshold set, the system would potentially be monitoring random noise, possibly confusing the user (see Fig. S7). This is a potential limitation of the method when simply using PCA, where the dynamic measurement with equilibrated $\Delta R/R$ responses at low concentrations has a large scatter surrounding the tight cluster of higher concentration responses (see Fig. 2C and Fig. 2D). By comparison the distribution of points within ANNs can clearly differentiate 4 vapors as well as the background nitrogen with tight clustering (see Fig. S8).

Additional information for classification of each analyte can be extracted from the characteristic time constants of each sensor and vapor combination. If the time constants differ for each sensor the temporal response within the first 3 principal components will become warped and no longer linear (see Fig. 3A). This is due to the faster sensors reaching their equilibrated $\Delta R/R$ response, while the slower sensors are at some smaller fraction of their equilibrated response (see Fig. S6). To use a sensor array within a complex and dynamic environment, incorporation of the sensor response time will provide the analyzer the ability to only pay attention to those sensors that would respond quickly, classifying a vapor immediately rather than waiting for all of the sensors to respond. With this sensor array the maximum wait time for a sensor to identify acetonitrile would take just over 6 min, as opposed to using a portion of the fast responding sensors to potentially classify the vapor within 2 min (Table S1). PSA is one of the faster sensors under these conditions and with analytes, it has separate monomers with a phenyl or hydroxymethyl bonded to the C—C backbone. The fastest equilibrating analyte to PSA sensor is methanol, having a response time of 1.77 min to reach 1 time constant. Factors that are likely contributing to the faster response time include the low steric hindrance from methanol compared to other vapors such as toluene that is large in comparison. The order of response to PSA is methanol < acetonitrile < dichloromethane < toluene, this corresponds to the trend of molecules with a lower molar refraction (R^2), a lower solute dipolarity-polarizability (p2H), a higher solute hydrogen bond acidity (a2H), a higher solute hydrogen bond basicity (b2H) and a lower solute hydrogen-gas-hexadecane partition coefficient (logL) (as defined in Table S3). Similar functional groups within sensors would be expected to respond with similar patterns, but with different intensities and time constants [10,21]. Where logL tends to be more positive for non-aqueous media and is highest in toluene, with toluene absorbing the slowest in PSA at 7 min and also with the highest variability in response time. Another contributing factor to the low response time of toluene is likely the a2H to bond with the alcohol group attached to the PSA backbone. R^2 relates to the polarizability of a mole of a substance, for a polymer like PSA with a hydrogen bonding alcohol group as well as the aromatic group. Less polarizable polymers such as PMVEMA where there are many more electronegative atoms present on the backbone reduce the overall polarizability of the polymer lowering the absorption of such vapors. In the case of PSA the b2H parameter does not appear to have a strong affect on the absorption of the vapor absorption rate [27]. The dependence appears to be larger for a2H where the higher the value the lower the response time. The odd solvent out when looking at this parameter is acetonitrile which responds faster than dichloromethane. Together these parameters indicate the importance of polar interactions in the absorption of vapors into PSA with some ability for hydrophobic interactions such as dispersion interactions. Similar descriptions of polymer solvent/vapor preferences can be concluded from Table S1 and Table S3. This characteristic behavior of each sensor and vapor combination will be useful within a system to be able to incorporate all the characteristic responses and behaviors. The best suited system is a ANN, capable of learning and adjusting to the slight changes that occur from a dynamic and complex environment. Plasticization from background (such as humidity) and self-plasticization has been demonstrated to have an effect on the time constants of the sensor array by increasing the rate of response [28]. This affect will vary from analyte to analyte and an approach for coping with such changes will need to be identified and incorporated within an ANN. With the additional changes in the overall pattern from the sensor array response this change in time constant may be able to be accounted for. In more extreme cases the effects from plasticizers might be significant. This may require more in depth training of the ANN using different plasticizers to minimize the altered sensor contributions to identification [28].

The ANN performance using sparse autoencoders for the compression of signals has resulted in the clear separation of 4 vapors tested (see Fig. 4). Normalization by this method is similar to the intensity factor, both are decreasing the variance associated with the concentration as well as the differing time constants between sensors. The calculated DBI values are indicative of an encoding of the sensor data within a space wherein different samples corresponding to the same vapor are less scattered, and samples drawn from different vapors exhibit a higher degree of separation (see Table 2). Gains are larger for the concentration ramping case, which is of particular interest given that this case is harder to cluster by traditional means as the absolute range of values changes as a function of concentration. The significant improvement in the DBI is indicative of the capacity for an encoding based on sparse filtering to compensate for this.

Many applications of machine olfaction are possible which require detection of a potentially large number of distinct vapors. While the preceding discussion has considered the unsupervised representation of olfactory signals in a neural network, it is also of

value to discuss this as a supervised classification problem, while potentially scaling up to a much larger number of sensors and vapors.

However with a significantly larger sensor array, with a greater diversity of sensors and more challenging application requirements, the current ANN architecture will be well equipped to encode response patterns to vapors with relative insensitivity to concentration while maximizing contrast between response profiles for different vapors. Moreover, concepts presented in additional discussion focused on ANNs provides a roadmap for addressing a wide array of applications in machine olfaction. This method has shown to be compatible with an array consisting of 8 chemically different sensors and 4 different vapors. Each vapor was separated into their own dense origin of principal component space, representing the unique pattern of each vapor. Among unsupervised approaches, sparse filtering has a tendency to cluster distinct vapors while also driving down variability attributed to concentration differences. For less controlled data it is also evident that a supervised approach may exhibit a significant degree of success provided vapor labels are known prior. A suitable example of a supervised approach is the combination of olfactory responses with those of machine vision, which could work in concert to identify an object *via* vision and relate the corresponding olfactory response. Vapor analysis results to date are reason to be optimistic as olfactory sensor data becomes more pervasive, and even more discriminative sensor arrays are produced. In considering challenges we have identified in this domain, convolutional neural networks and recurrent neural networks (e.g. LSTM) are likely to be invaluable tools in scaling up the sophistication of olfactory sensing systems in much the same way that this has transpired for efforts in machine vision.

For a system that implements machine olfaction on a longer time course, there is also the need to consider changes to sensor characteristics over time, and replacement of sensors. Sparsity is valuable in supporting this, as any individual unit is likely to play a partial role in only a fraction of classification decisions. With respect to strategies for training neural networks, one common practice that has emerged in this area is known as dropout. This is a strategy commonly employed to prevent learning weights that overfit any training data. The process of dropout is such that a random subset of units at any layer of the neural network are excluded from carrying signals forward or adjustment of weights during training. This prevents trained units from becoming co-dependent. This is likely also to guard against sudden changes in the properties of any single unit within the network.

In the ideal case, the weights of the network may be adjusted over time to accommodate for gradual or sudden changes to sensor properties due to poisoning, malfunction or chemical breakdown. One way of dealing with these issues is by re-calibration of network weights through exposure to a representative sample of known vapors. However, a solution that adjusts behavior online may be more desirable and practical. Assuming that network output is relatively insensitive to sudden changes to the output of a single sensor, and temporal degradation of sensor output is sufficiently slow, it is reasonable to consider gradual online re-calibration based on classification outputs. There are many ways this might be implemented *via* strategies such as the novelty rule [29] or pseudo-recurrent networks [30]. One relatively simple strategy is to include a sigmoidal non-linearity at the output while treating the difference of the largest output above threshold from 1 as an error signal. This assumes that the network has been initialized to a relatively stable state, and will also benefit from the properties produced by sparse coding and dropout. The strategies for solving this problem closely mirror those for addressing the problem of catastrophic interference [31,32] and requires considering both sensor properties, and network properties.

5. Conclusions

We have presented in this paper the development and foundation for a new method of vapor identification to grow. Using the combined intensities of the sensor array to normalize the changing responses eases the burden on the system and the user to accurately identify a vapor. This opens the door to creating segregated areas within principal component space as well as component space of the ANNs for an "on or off" approach to vapor identification around a threshold of detection. With a base ANN set, increased complexity and added information to the system with help with the temporal identification and quantification of different vapors. Future developments could also include other sensory inputs, such as machine vision to aid in the supervised identification of different vapors and their sources.

Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada (Grant No. 314025), the Canada Foundation for Innovation (CFI) Grant No.: 315751, Manitoba Research and Innovation Fund and the University of Manitoba for their financial support. Additional funding was provided in part from the Canada Research Chairs (CRC) Grant No.: 301917 Program for SR and MSF.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.snb.2016.03.059.

References

[1] R. Gutierrez-Osuna, Pattern analysis for machine olfaction: a review, IEEE Sens. J. 2 (2002) 189–202.
[2] B. Malnic, J. Hirono, T. Sato, L. Buck, Combinatorial receptor codes for odors, Cell 96 (1999) 713–723.
[3] B. Ache, J. Young, Olfaction: diverse species, conserved principles, Neuron 48 (2005) 417–430.
[4] J. Lundström, S. Boesveldt, J. Albrecht, Central processing the chemical senses: an overview, ACS Chem. Neurosci. 2 (2011) 5–16.
[5] M. Schumuker, G. Schneider, Processing and classification of chemical data inspired by insect olfaction, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 20285–20289.
[6] A. Karpathy, L. Fei-Fei, Deep visual alignments for generating image descriptions, Proc IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR (2015).
[7] W.S. MucCulloch, W. Pitts, A logical calculus of the ideas immanent in the nervious activity, Bull. Math Biophys. 5 (1943) 115–133.
[8] J.J. Hopfield, Olfactory computation and object perception, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 6462–6466.
[9] M.S. Freund, N.S. Lewis, A chemically diverse conducting polymer-based electronic nose, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 2652–2656.
[10] N. Lewis, Comparisons between mammalian and artificial olfaction based on arrays of carbon black-polymer composite vapor detectors, Acc. Chem. Res. 37 (2004) 663–672.
[11] E.J. Severin, J.D. Brett, N. Lewis, An investigation of the concentration dependance and response to analayte mixtures of carbon black/insulating organic polymer composite vapor detectors, Anal. Chem. 72 (2000) 658–668.
[12] J.D. Brett, E.J. Severin, N. Lewis, Trends in odor inensity for human and electronic noses: relative roles of odorant vapor pressure vs. molecularly specific odorant binding, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 5442–5447.
[13] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data iwht neural networks, Science 313 (2006) 504–507.
[14] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy emplyed by V1? Vision Res. 37 (1997) 3311–3325.
[15] A. Coates, A.Y. Ng, H. Lee, An analysis of single-layer networks in unsupervised feature learning, Int. Conf. Artif. Intell. Stat. (2011) 215–223.
[16] M. Bhand, R. Mudur, B. Suresh, A. Saxe, A.Y. Ng, Unsupervised learning models of primary cortical receptive fields and receptive field plasticity, Adv. Neural Inf. Process. Syst. (2011) 1971–1979.

[17] J. Ngiam, P. Wei Koh, Z. Chen, S. Nhaskar, A.Y. Ng, Sparse filtering, Adv. Neural Inf. Process. Syst. 24 (2011) 1125–1133.
[18] M. Carandini, D.J. Heeger, Normlaization as a canonical neural computation, Nat. Rev. Neurosci. 13 (2011) 51–62.
[19] M.R. Kumar, S. Ryman, M.O. Tareq, D. Buchanan, M.S. Freund, Chemical diversity in electrochemically deposited conducting polymer-based sensor arrays, Sens. Actuators B Chem. 202 (2014) 600–608.
[20] M.E. Hossain, G.M.A. Rahman, M.S. Freund, D.S. Jayas, N.D.G. White, C. Shafai, et al., Fabrication and optimization of a conducting polymer sensor array using stored grain model volatiles, J. Agric. Food Chem. 60 (2012) 2863–2873.
[21] M.H. Abraham, Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes, Chem. Rev. 96 (1992) 73–83.
[22] J.W. Grate, M.H. Abraham, Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays, Sens. Actuators B Chem. 3 (1991) 85–111.
[23] S.M. Briglin, N. Lewis, Characterization of the temporal response profile of carbon black −Polymer composite detectors to volatile organic vapors, J. Phys. Chem. B 107 (2003) 11031–11042.
[24] J. Monroy, J. González-Jiménez, J. Blanco, Overcoming the slow recovery of MOX gas sensors through a system modeling approach, Sensors 12 (2012) 13664–13680.
[25] T.A. Cleland, et al., Relational representation in the olfactory system, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1953–1958.
[26] S. Hochreiter, J. Schimduber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[27] P. Sakellariou, M.H. Abraham, G.S. Whiting, Solubility of poly(ethylene oxide): effect of molecular weight, end groups and temperature, Colloid Polym. Sci. 272 (1994) 872–875.
[28] M.E. Koscho, R.H. Grubbs, N.S. Lewis, Properties of vapor detector arrays fromed through plasticization of carbon black—organic polymer composites, Anal. Chem. 74 (2002) 1307–1315.
[29] C.A. Kortge, Episodic memory in connectionist networks, in: Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, Hillsdale, NJL Lawrence Erlbaum, 1990, pp. 764–771.
[30] J. McClelland, B. McNaughton, R. O'Reilly, Why there are complementary learning systems in the hippocacmpus and neocortex: insights form the successes and failures of connectionist models of learning and memory, Psychol. Rev. 102 (1995) 419–457.
[31] M. McCloskey, N. Cohen, Catastrophic interference in connectionist networks: the sequential learning Problem, in: G.H. Bower (Ed.), Psychol. Learn. Motiv., 24, Academic Press, NY, 1989, pp. 109–164.
[32] R. Ratcliff, Connectionist models of recognition memory: constraints imposed by learning and forgetting functions, Psychol. Rev. 97 (1990) 285–308.

Biographies

Shaun K. Ryman received his BS degree in chemistry at the University of Manitoba in 2013. He is currently a MS student at the Univeristy of Manitoba in the Department of Chemistry. His research work involves using statistical methods to evaluate and improve the temporal responses of gas sensor arrays for machine olfaction.

Neil D.B. Bruce is currently an Assistant Professor in the Department of Computer Science at the University of Manitoba, Canada. He has been a postdoctoral fellow at the Centre for Vision Research at York University, and at INRIA Sophia Antipolis in France. He holds a Ph.D. in Computer Science (York University, 2008), M.A.Sc in System Design Engineering (University of Waterloo, 2003), and an Honours B.Sc. with Majors in Computer Science and Mathematics (University of Guelph, 2001).

M.S. Freund received a B.S. Degree in Chemistry from Florida Atlantic University in 1987 and a Ph.D. in Analytical Chemistry in 1992 from the University of Florida. He currently holds the positions of Professor and Head of the Department of Chemistry at Florida Institute of Technology.