"Super E-Noses": Multi-Layer Perceptron Classification of Volatile Odorants from the Firing Rates of Cross-Species Olfactory Receptor Arrays

Luqman R. Bachtiar, Charles P. Unsworth, Member, IEEE, Richard D. Newcomb

Abstract— **Current electronic noses, or e-noses, that employ insect odorant receptors (Ors) as their sensory front end are potentially limited by the fact that the Ors come from a single species. In addition, a realistic e-nose also demands low numbers of Ors at its sensory front end due to the difficulties of receptor/sensor integration and functionalisation.**

In this work, we report the first investigations of a 'Super E-Nose' that incorporates Ors from both the model organism *Drosophila melanogaster* fruit fly (DmOr) and the mosquito, *Anopheles gambiae* (AgOr). Furthermore, we report how an Artificial Neural Network (ANN), in the form of a hybrid double hidden layer Multi-Layer Perceptron (MLP), can be used to determine the optimal Ors that provide the best prediction performance in the classification of unknown odorants into their respective chemical class.

Our findings demonstrate how 3-Or arrays consisting of DmOr only, AgOr only, or cross-species DmOr-AgOr combinations correctly classified all unknown odorants of the validation set. In addition, we report that all 3-Or combinations perform equally well as the complete 74 DmOr-AgOr array. Thus, the results of this work support further investigation into cross-species 'Super E-noses' coupled with hybrid MLPs for the classification of unknown odorants.

I. Introduction

An electronic nose or e-nose is a device used to detect odors and flavors. E-noses have an assortment of applications such as determining ripeness of fruit, detection of illegal contraband, and in medical diagnosis [1]. Such devices typically consist of an aroma delivery system, a chamber housing an array of sensors, a processing system for converting the chemical signals into a digital electrical signal, and a computer microprocessor [1-3]. The sensory array is an important component of the device as it must respond to a wide range of chemical classes and discriminate mixtures of possible chemicals. As insects are highly reliant on their olfactory senses for feeding, mating, and other behaviors [4], an e-nose that incorporates insect olfactory receptors could be a valuable alternative to the human nose, which is traditionally used.

This work was supported by New Zealand's New Economy Research Fund C06X0701.

L. R. Bachtiar is with the Department of Engineering Science at The University of Auckland, Auckland 1010, New Zealand. (phone: +64-9-373-7599 x87061; e-mail: l.bachtiar@auckland.ac.nz).

C. P. Unsworth is a Senior Lecturer (and Main Supervisor) at the Department of Engineering Science at The University of Auckland, Auckland 1010, New Zealand (e-mail: c.unsworth@auckland.ac.nz).

R. D. Newcomb is with The New Zealand Institute of Plant & Food Research, Private Bag 92169, Auckland 1142, New Zealand (e-mail: Richard.Newcomb@plantandfood.co.nz).

Insects use their olfactory organs, the antenna and maxillary palp (Fig. 1), to identify a myriad of chemicals [2, 4]. Covering the antennae and maxillary palps are hair-like sensilla which house Olfactory Receptor Neurons (ORNs) that encode qualitative, quantitative, and temporal information about odors. Odorant detection begins when airborne odors enter pores along the sensilla. Within the aqueous environment inside the sensilla, there are many odorant-binding proteins that solubilize and transport the odorants to receptors on dendrites of ORNs which extend into the lymph. Typical resting potential of ORNs is approximately -65mV [5] and with the activation of an appropriate odorant, an influx of Na^+ and Ca^{2+} results in a positive (less negative) spike in membrane potential by about 20mV. Such a change in membrane potential induces the propagation of an action potential to the antennal lobe, where input is modulated by interneurons and transmitted by projection neurons to higher brain regions for processing.

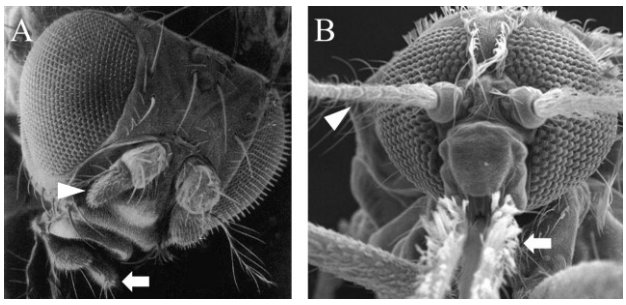

Figure 1. Electron micrograph of the A) *Drosophila melanogaster* [6] and B) *Anopheles gambiae* [7] antenna (arrowhead) and maxillary palp (arrow).

Recording of spontaneous insect receptor responses to odors is made possible through the use of recombinant gene technology in which Odorant Receptors (Ors) are ectopically expressed in a mutant 'empty' *Drosophila melanogastser* ORN [8, 9]. A combinatorial model of odor coding [8] has been found in *D. melanogaster* [8] and in *Anopheles gambiae* [9] in which individual receptors responded to subsets of odorants and individual odorants activated subsets of receptors. In this work, we use Artificial Neural Networks (ANNs) in the form of a Multi-Layer Perceptron (MLP) as a signal processing backend of an e-nose to classify odorants by analyzing the information embedded in the neuronal firing rates of insect Ors. ANNs are a powerful tool for extracting essential features from data sets. The study of olfactory recognition presents highly complex relationships between the data and the classes they belong to and ANNs are a viable method used to predict and classify unknown or unseen odorants.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

II. METHOD

A. Data Used in the Study

In this work, we utilize the spontaneous Or response data of the *D. melanogaster* obtained from the work of Hallem [8] and the *A. gambiae* spontaneous responses recorded by Carey [9]; spontaneous activity is quantified from the number of spikes recorded in 1 second of spontaneous activity [10]. Or recordings in both studies were obtained using a mutant ORN of *D. melanogaster* [8, 9]. The studies conducted by Hallem [8] and Carey [9] recorded Or responses to 108 and 110 odorants respectively. To investigate the application of a cross-species sensor, we identified the common odorants used in both studies. Consequently, the data used in this work consist of the firing rates of 24 *D. melanogaster* odorant receptors (DmOrs) and 50 *A. gambiae* odorant receptors (AgOrs) in response to 34 common odorants. These odorants fell into 5 distinct chemical classes (with the number of chemicals listed in parentheses): Acid (7); Ketone (5); Aromatic (6); Alcohol (9) and Ester (7). To enhance network learning [11], the data were preprocessed by zero-mean and normalization prior to feeding into the MLP system.

B. Selection of ANN Architecture and Training Scheme

In this work, we employ the Artificial Neural Network (ANN) architecture of a feed forward Multi-Layer Perceptron (MLP) [12-15]. Fig. 2 presents a simplified schematic of a double hidden layer hybrid MLP system. The odor response of the ORN array to an odorant is used as the input vector, which is fed into the input layer and subsequent layers. Through network training, the MLP system is able to provide generalizations of the data used [16]. This is achieved by using a supervised back-propagation learning scheme in which the weighting functions are adjusted at each time-step (epoch) to produce a progressively effective network. Back-propagation is known to perform ineffectively when a local minimum is present [17], thus a momentum function was applied to the system to address this issue [18].

Weighting functions of the network are represented by the hexagons in Fig. 2, and their initial values obtained from a symmetric Gaussian distribution with a zero mean and variance of unity. Small weighting functions were chosen to ensure optimization of the 'weight' decay regularizer, which enhances network generalization, and to prevent over-fitting of the MLP systems [12]. Together with a binary sigmoidal function at each layer, the weighting functions produce a network output: an output close to 1 was used to denote identification of a correct chemical class whilst an output close to 0 represents an incorrect class.

To quantify the degree of network learning, a validation set is applied to the trained MLP system. This set is composed of a random set of 5 chemical odorant vectors collected from the 34 available odorants. For each sample of bootstrapping, the validation set was obtained by randomly selecting 10% of odorants from each chemical class. The numbers of odorants of each chemical class in the validation set were: Acid (1); Ketone (1); Aromatic (1); Alcohol (1) and Ester (1). The remaining 29 odorants were subsequently used for the corresponding bootstrapping training set sample. Final results of an MLP performance were obtained through bootstrapping methods so as to portray the variability of simulations and to provide more accurate estimates of predicted values [12, 13, 19]. We chose to repeatedly sample 10,000 training and validation sets from the raw data for retraining and validation of the system [19-21].

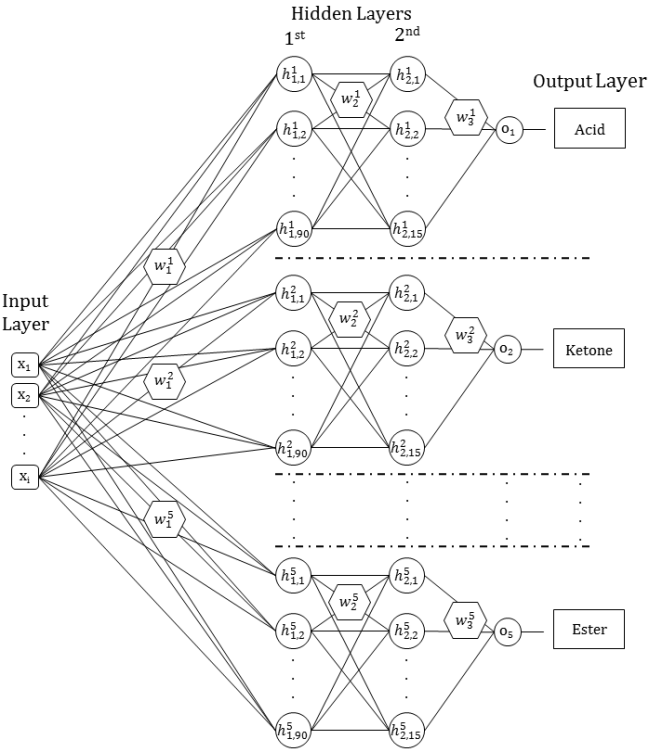

Figure 2. Schematic of a double hidden layer hybrid Multi-Layer Perceptron (MLP) used to class odorants into their chemical classes.

Preliminary tests were performed wherein network learning ended based on fixed epochs or when the prediction error fell below a defined tolerance [22]. We found 200 epochs presented satisfactory CPU run times and a satisfactory error prediction falling below the desired 0.01 threshold.

We have previously discovered that the quality of odorant classification relies heavily on a network's size and structure [12, 13] as it inherently affects the network's critical learning time and generalization capabilities [16]. By alternating the number of hidden layers used, the number of hidden neurons of the layer(s), and the MLP configuration (single or hybrid system), we were able to investigate the effect of using different MLP architectures and their subsequent classification performances. A hybrid MLP system is a set of independent single-output MLPs working in parallel and the number of MLPs in the series is determined by the number of classes present in the data set; in this work, 5 MLPs are used in the hybrid MLP system. Although investigating the different MLP architectures was a long and demanding process, determining the optimal MLP system is important due to its capability of extracting higher-order information from a given data set to provide an improved classification performance [23].

The performance difference between MLP architectures was based on the highest prediction accuracy of the validation odorant with the lowest prediction across the validation set. These values were obtained using *k*-fold cross-validation [24, 25], which allowed for an efficient simulation time while providing an improved accuracy of measuring classifier performance over the fixed validation method [12-15]. The results presented in this work were obtained from the optimal MLP architecture (Fig. 2).

C. *Analyzing Insect Odorant Receptor Combinations*

Due to the long and complicated process of developing and recording insect odorant receptors *in vitro*, it is desirable to have an e-nose that operates effectively with a low number of Ors. We investigate the changes in MLP performance when using a minimal set of Ors and propose the identification and use of an optimal combination of 3 Ors from the available 74 Ors. A bank consisting of all 64,824 combinations was tested with the optimal MLP system. Different sequences of the same combination are not included and selection of a receptor does not occur twice in a given combination [12]. Similar to assessing the optimal MLP architecture, *k*-fold cross-validation was used to compare the performance between Or combinations and evaluated by the highest prediction accuracy of the validation odorant with the lowest prediction across the validation set.

III. RESULTS

From preliminary tests we found an optimal MLP architecture that produced the best performance: a double hidden layer hybrid MLP system with 90 neurons in the first hidden layer and 15 neurons in the second hidden layer (Fig. 2). Such an MLP system is known to capture more of the complex relationships present in the data [12, 13, 15, 26] allowing for successful classification of unknown odorants.

A threshold value was applied to quantify MLP learning. This value was determined by calculating the probability of choosing the correct class of a validation vector: 1/5 or 20%. With a 5% added safeguard, the 25% conservative threshold value was defined; the dotted horizontal line in Fig. 3 illustrates this conservative threshold value. Successful classification is established when the calculated prediction of the validation odorant exceeds the conservative threshold value.

In Fig. 3, we present the classification performance using the complete 74 DmOr-Agor array, the best 3 DmOr (DmOr85f, DmOr88a and DmOr98a), the best 3 AgOr (AgOr73, AgOr75 and AgOr76) and the best 3 DmOr-AgOr (DmOr9a, DmOr22a and AgOr67) combinations. We found that all 4 arrays successfully classified the validation vectors of all the chemical classes, surpassing the threshold value, including the worst possible case (lower error bar). Slight variation in mean prediction (%) of the validation vectors was observed between the 3 Or combinations (the higher prediction combination in brackets): Acid (DmOr), Ketone (DmOr and DmOr-AgOr), Aromatic (AgOr), Alcohol (DmOr and DmOr-AgOr) and Ester (DmOr-AgOr). Furthermore, the average prediction across the validation set was 64.6% for the 3 DmOr, 63.2% for the 3 AgOr and 67.9% for the 3 DmOr-AgOr combination. Hence, the best performing 3-Or combination was found to consist of 2 *D. melanogaster* odorant receptors and 1 *A. gambiae* odorant receptor: DmOr22a, AgOr67 and DmOr9a.

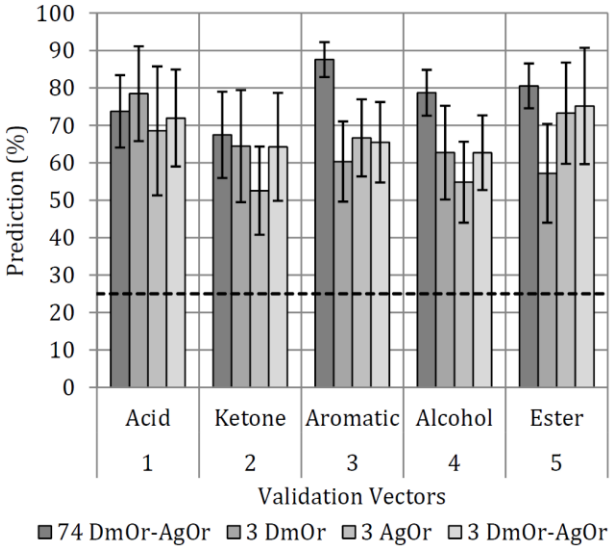

Figure 3. Performance of the optimal hybrid MLP system when using the complete 74 DmOr-AgOr array, best 3 DmOr, best 3 AgOr and best 3 DmOr-AgOr combination. The horizontal broken lines represent the 25% conservative threshold value that defines correct classification of an odorant of the validation set. Bootstrapping results comprise of the number of odorants on average surpassing the treshold, the worst possible (lower error bar) and best possible (upper error bar) case of classification.

IV. CONCLUSION

In this exploratory work, we used an Artificial Neural Network (ANN) in the form of a double hidden layer hybrid Multi-Layer Perceptron (MLP) for the classification of odorants based on the firing rates of *Drosophila melanogaster* odorant receptors (DmOrs) and *Anopheles gambiae* odorant receptors (AgOrs). Due to the long and complicated process of expressing and functionally characterizing insect odorant receptors (Ors) *in vitro*, we sought to use a minimal combination of Ors for ANN analysis. MLP classification using an optimal 3 Or combination, regardless of species, was found to perform equally well in predicting all 5 odorant chemical classes as the complete 74 DmOr-AgOr array. The 3 Or combinations tested were DmOr only, AgOr only, and a DmOr-AgOr mix, and all correctly identified all validation vectors. The best performing 3-Or combination was found to consist of 2 *D. melanogaster* odorant receptors and 1 *A. gambiae* odorant receptor: DmOr22a, AgOr67 and DmOr9a. Thus, the results demonstrate for the first time the viability of implementing cross-species Ors as the sensory processing system of a "Super E-nose". We hope that these results will pave the way for future developments of 'Super E-Nose Technology' utilizing additional Or species, such as the *Apis mellifera* (honeybee) or *Utetheisa ornatrix* (polyphemus moth), and the creation of multi-species Or arrays.

ACKNOWLEDGEMENTS

We would like to thank John Carlson and Allison Carey from the Department of Molecular, Cellular and Developmental Biology, Yale University, for kindly providing the data and their helpful discussions in the preparation of this paper.

REFERENCES

[1] J. W. Gardner and P. N. Bartlett, *Electronic Noses: Principles and Applications.* Oxford University Press Oxford, 1999.

[2] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensors and Actuators B: Chemical,* vol. 18, pp. 210-211, 1994.

[3] B. Kowalski and C. Bender, "Pattern recognition. Powerful approach to interpreting chemical data," *J. Am. Chem. Soc.,* vol. 94, pp. 5632-5639, 1972.

[4] W. Takken, "The Role of Olfaction in Host-Seeking of Mosquitos - a Review," *Insect Sci Appl,* vol. 12, pp. 287-295, Jun, 1991.

[5] S. Firestein, "How the olfactory system makes sense of scents," *Nature,* vol. 413, pp. 211-218, 2001.

[6] J. Riesgo-Escovar, D. Raha and J. R. Carlson, "Requirement for a phospholipase C in odor response: overlap between olfaction and vision in Drosophila," *Proc. Natl. Acad. Sci. U. S. A.,* vol. 92, pp. 2864-2868, Mar 28, 1995.

[7] (2006, April 20). *Head region of Anopheles gambiae*. Available: http://phil.cdc.gov/PHIL_Images/8780/8780_lores.jpg.

[8] E. A. Hallem and J. R. Carlson, "Coding of Odors by a Receptor Repertoire," *Cell,* vol. 125, pp. 143-160, 2006.

[9] A. F. Carey, G. Wang, C. Su, L. J. Zwiebel and J. R. Carlson, "Odorant reception in the malaria mosquito Anopheles gambiae," *Nature,* vol. 464, pp. 66-71, 2010.

[10] E. A. Hallem, M. G. Ho and J. R. Carlson, "The Molecular Basis of Odor Coding in the Drosophila Antenna," *Cell,* vol. 117, pp. 965-979, 2004.

[11] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *Nuclear Science, IEEE Transactions On,* vol. 44, pp. 1464-1468, 1997.

[12] L. R. Bachtiar, C. P. Unsworth, R. D. Newcomb and E. J. Crampin, "Multilayer perceptron classification of unknown volatile chemicals from the firing rates of insect olfactory sensory neurons and its application to biosensor design," *Neural Comput.,* vol. 25, pp. 259-287, 2013.

[13] L. R. Bachtiar, C. P. Unsworth and R. D. Newcomb, "Application of artificial neural networks on mosquito olfactory receptor neurons for an olfactory biosensor," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE,* 2013, pp. 5390-5393.

[14] L. R. Bachtiar, C. P. Unsworth, R. D. Newcomb and E. J. Crampin, "Predicting odorant chemical class from odorant descriptor values with an assembly of multi-layer perceptrons," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE,* 2011, pp. 2756-2759.

[15] L. R. Bachtiar, C. P. Unsworth, R. D. Newcomb and E. J. Crampin, "Using artificial neural networks to classify unknown volatile chemicals from the firings of insect olfactory sensory neurons," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE,* 2011, pp. 2752-2755.

[16] E. Baum and D. Haussler, "What size net gives valid generalization?" *Neural Computation,* vol. 1, pp. 151-160, 1989.

[17] P. S. Churchland and T. J. Sejnowski, *The Computational Brain.* Cambridge, Mass.: MIT Press, 1992.

[18] S. Huang, K. K. Tan and K. Z. Tang, *Neural Network Control: Theory and Applications.* Baldock, Hertfordshire, England: Research Studies Press, 2004.

[19] R. Tibshirani, "A comparison of some error estimates for neural network models," *Neural Computation,* vol. 8, pp. 152-163, 1996.

[20] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics,* vol. 35, pp. 352-359, 2002.

[21] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap.* Chapman & Hall/CRC, 1993.

[22] C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford university press, 1995.

[23] I. D. Longstaff and J. F. Cross, "A pattern recognition approach to understanding the multi-layer perception," *Pattern Recognition Letters,* vol. 5, pp. 315-319, 1987.

[24] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques.* Morgan kaufmann, 2006.

[25] A. Blum, A. Kalai and J. Langford, "Beating the hold-out: Bounds for k-fold and progressive cross-validation," in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory,* 1999, pp. 203-208.

[26] J. W. Gardner, E. L. Hines and M. Wilkinson, "Application of Artificial Neural Networks to an Electronic Olfactory System," *Measurement Science and Technology,* vol. 1, pp. 446, 1990.