# Limitations in odour recognition and generalisation in a neuromorphic olfactory circuit

Nik Dennler[1,2], André van Schaik[2] and Michael Schmuker[1*]

[1]Biocomputation Group, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, United Kingdom.
[2]International Centre for Neuromorphic Systems, The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Kingswood, NSW 2747, Australia.

*Corresponding author(s). E-mail(s): m.schmuker@herts.ac.uk;
Contributing authors: n.dennler2@herts.ac.uk;
a.vanschaik@westernsydney.edu.au;

Arising from Imam & Cleland [1].

Neuromorphic computing is one of the few current approaches that have the potential to significantly reduce power consumption in Machine Learning and Artificial Intelligence, and has drawn vast inspiration from considerations of biological systems and circuits. In their work, Imam & Cleland presented a neuromorphic odour-learning algorithm that is inspired by mammalian olfactory bulb circuitry, which they assessed by considering its performance in "rapid online learning and identification" of gaseous odorants and odorless gases (short "gases") using a set of gas sensor recordings of different odour presentations and corrupting them by impulse noise. We replicated parts of the study and discovered limitations thereof, which are 1) that the dataset used suffers from sensor drift and a non-randomised measurement protocol that render it of limited use for odour identification benchmarks, and 2) that the model is restricted in its ability to generalise over repeated presentations of the same gas. Therefore, a validation of the model that goes beyond restoring a previously learned data sample remains to be shown, in particular its coherence

with the attributed capabilities of robustness and broad generalisation beyond experience, as well as its suitability to realistic odour identification tasks.

The first limitation of the study relates to restrictions in the dataset used to validate the olfactory bulb network. The dataset consisted of recordings from 72 Metal Oxide (MOx) gas sensors mounted in a wind tunnel [2] (Fig. 1a). MOx sensors are highly prone to sensor drift, causing short- and long-term fluctuations in the sensors' baseline conductance and their responsiveness [3]. The most effective way of tackling the effect of drift on sensor response is to randomise the gas presentations over time during the recording. The dataset used here does not have this property: Recordings were acquired in gas-specific batches over the course of nine months (Fig. 1b). The non-randomness, together with the dominating presence of sensor drift contaminations, allows for successful gas classification *before* the gas is presented, which renders this dataset largely unsuitable for classification tasks [4]. Drift contaminations could be partly mitigated by subtracting the baseline, i.e., the sensor response right before gas exposure, which is a widely used procedure when using MOx sensor data [5]. No baseline subtraction was performed in the discussed study, suggesting that the reported findings about odour learning and recognition may be skewed, and potentially invalid, due to the limitations of the dataset.

We repeated the simulations for a range of conditions, using the publicly available code by the authors. As in the original work, the model was trained on 10 gases, and tested on 10 occlusions for each gas, i.e., 100 samples total. If not otherwise stated, 60% of the data was occluded by impulse noise when testing. We successfully replicated the authors' Jaccard similarity coefficient plot ([1], Fig. 4b), using the same raw data points for composing training and test sets, and sampling the recordings at $t = 90s$ (Fig. 1a, 2a). The result appears to demonstrate robust recognition of the Toluene gas instance. Paradoxically, the same level of "recognition" of Toluene can be obtained in the absence of gas, using samples obtained at $t = 15s$, *before* the release of odour into the wind tunnel at $t = 20s$ (Fig. 1a, 2b). Therefore, the high Jaccard score for Toluene should be considered an artefact of sensor drift, and is unsuitable to substantiate a capability of the model to recognise odours [4].

In addition to the dataset's limitations, we found restrictions in the model's capability to generalise over different recordings of the same stimulus. Generalisation is an important property of any pattern recognition system [6]. The authors convincingly show that the model can restore input patterns corrupted by impulse noise. However, in most instances the authors tested recognition on the same sample that was used for training, occluding 60% of the sample with noise. 40% of each training sample were present unchanged in the corresponding testing sample. By using overlapping portions of data for both training and

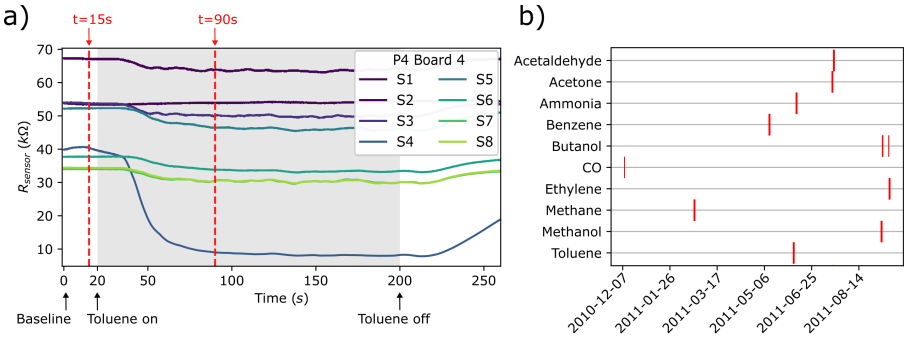


**Fig. 1**: Overview of the data collection protocol. **a)** Example sensor resistance measurement for all sensors on one sensor board (location P4, module 4, Toluene, 0.21 m s$^{-1}$ airflow velocity, 5 V operating voltage, trial 2). The shaded area denotes the period during which the gas is injected into the wind tunnel, the red dotted lines indicate where the data is sampled in [1] ($t = 90$ s) and in this work ($t = 15$ s and $t = 90$ s). **b)** Timestamps of the gas sensor recordings from which, in [1], one trial per gas was sampled. Each vertical line represents multiple trials (up to 20), which were performed too close to each other for them to be visually distinguishable in this representation.

.

testing, the statistical independence required for a robust assessment of generalisation is not maintained. A real odorant recognition and signal restoration system would rarely encounter the exact same stimulus twice, once in a clean and once in a corrupted version. Therefore, assessing the model's capability to recognise and restore patterns from separate recordings is essential to judge its relevance.

For most gas and parameter combinations, the dataset contains 20 repetitions. We repeated the experiment using separate repetitions for training and testing and found that gas identity could not be recognised in occluded samples (Fig. 2c). Recognition scores were further reduced when subtracting the baseline from training and testing data (Fig. 2d). In this configuration, aimed at mitigating sensor drift, recognition across repetitions failed even for samples without any noise occlusion (Fig. 2e).

Finally, we note the authors' claim of superior performance of the EPL network over advanced machine learning methods, derived from comparisons with multilayer denoising autoencoders and other methods. We wonder if this can be upheld in its generality, and tested whether the same performance could be reached with much simpler methods. In fact, we found that the evaluation task demonstrated in Fig 4. of the study is simple enough to be effectively addressed by using a toy model consisting of a simple hash table, such as a *Python* dictionary, without the need for complex machine learning techniques. By hashing

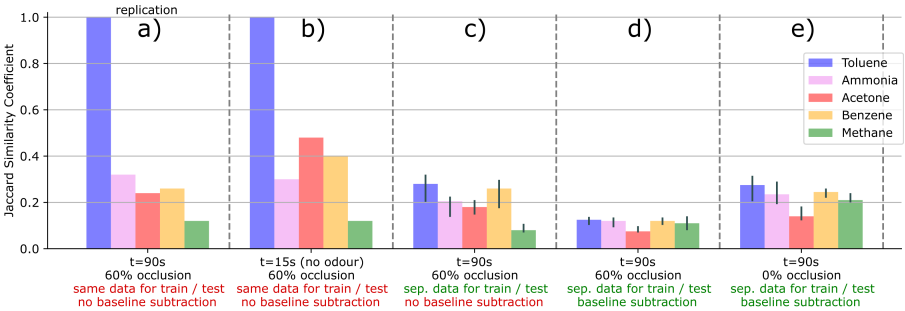


**Fig. 2**: Ablation study, evaluating the EPL network under a range of conditions. a) Jaccard similarity coefficient of the networks response to occluded Toluene and the learned odour representations, after five successive gamma cycles. Replicated from [1]. b) Spurious recognition in absence of gas in the wind tunnel. c,d,e) Recognition failure on different repetitions, c) without and d) with baseline subtraction, and e) without sample occlusion. Bar heights and error bars respectively display the median and interquartile range of the algorithms' output similarity, as resulting from 72 sensor measurements across 10 Toluene representations. Only five out of 10 gases are depicted for clarity.

the training samples (i.e. one for each class, as in the original paper), and then computing and ranking the amount of overlap between a test sample and the stored training samples, one can estimate the most likely class, from which the similarity to the actual class can be computed (see Alg. 1&2). We acknowledge that this approach is neither expected to be robust to stochastic noise or systematic variances, nor suitable for any realistic odour recognition task. Yet, it matches or surpasses the EPL network in recognition accuracy on the particular task under the original and the extended conditions discussed (Fig. S1), and outperforms it in runtime (Fig. S2). Equivalent and exceeding results can be achieved for similar tasks on this dataset using support vector machines [4]. While this does not disprove that the proposed model has denoising, reconstructing, and recognition capabilities, it demonstrates that the evaluation approach used is insufficient to support the favourable comparison to advanced machine learning algorithms.

We conclude that the capability of the proposed model to identify learned odourants appears to be limited to corrupted versions of the training data. It failed to generalise to data outside the training set: Repetitions of learned gases were not recognised if these repetitions were not part of the training data. Imam & Cleland's model is an elegant example for an implementation of a biologically plausible model on neuromorphic hardware that can restore learned signals corrupted by noise. However, due to the restricted generalisation capability of the model, and to the limitations of the data used, it

cannot be claimed that it solves the problem of odour learning and identification under a realistic scenario. We hope that raising awareness about these limitations paves the way towards improved neuromorphic models for robust gas recognition that can solve real-world odour recognition tasks.

## Data Availability

The data used for our experiments are publicly available and described in Vergara *et al.* [2].

## Code Availability

Our experiments were based on the code released by the authors with the original study. Our adaptations and instructions for replication, together with the data, are available at https://github.com/BioMachineLearning/EPLNetwork_ImamCleland2020.

## Author Contribution Statement

**Nik Dennler:** Conceptualisation, Investigation, Formal Analysis, Software, Visualisation, Writing – Original Draft, Writing – Review & Editing. **André van Schaik:** Conceptualisation, Writing – Review & Editing, Supervision. **Michael Schmuker:** Conceptualisation, Writing – Review & Editing, Funding acquisition, Supervision.

## Declaration of Competing Interest

The authors declare no competing interests.

## Acknowledgements

We thank N. Imam and T.A. Cleland for their valuable feedback and suggestions. Further, we thank D. Drix, M. Psarrou, S. Rastogi and S. Sutton for fruitful discussions. M.S. was funded from EU H2020 Grant Human Brain Project SGA3 (#945539). This project is supported by the NSF/CIHR/DFG/FRQ/UKRI-MRC Next Generation Networks for Neuroscience Program (NSF #2014217 / MRC #MR/T046759/1 "Odor2Action").

## References

[1] Imam, N., Cleland, T.A.: Rapid online learning and robust recall in a neuromorphic olfactory circuit. Nature Machine Intelligence **2**(3), 181–191 (2020). https://doi.org/10.1038/s42256-020-0159-4

[2] Vergara, A., Fonollosa, J., Mahiques, J., Trincavelli, M., Rulkov, N., Huerta, R.: On the performance of gas sensor arrays in open sampling systems using inhibitory support vector machines. Sensors and Actuators B: Chemical **185**, 462–477 (2013). https://doi.org/10.1016/j.snb.2013.05.027

[3] Vergara, A., Vembu, S., Ayhan, T., Ryan, M.A., Homer, M.L., Huerta, R.: Chemical gas sensor drift compensation using classifier ensembles. Sensors and Actuators B: Chemical **166-167**, 320–329 (2012). https://doi.org/10.1016/j.snb.2012.01.074

[4] Dennler, N., Rastogi, S., Fonollosa, J., van Schaik, A., Schmuker, M.: Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks. Sensors and Actuators B: Chemical **361**, 131668 (2022). https://doi.org/10.1016/j.snb.2022.131668

[5] Hines, E.L., Llobet, E., Gardner, J.: Electronic noses: a review of signal processing techniques. IEE Proceedings-Circuits, Devices and Systems **146**(6), 297–310 (1999)

[6] Gareth, J., Daniela, W., Trevor, H., Robert, T.: An Introduction to Statistical Learning: with Applications in R. Springer, New York, NY (2013). https://doi.org/10.1007/978-1-4614-7138-7

# Supplementary Information

## Hash Table Algorithm

```
learned_odour_representations = {idx: odour for idx, odour in
    enumerate(training_data)}
```

**Algorithm 1**: Training / One-shot learning (*Python* code). For each training sample, fill dictionary with odour key and representation.

```
denoised_testing_data = {}
for idx_test, testing_odour in enumerate(testing_data):
    best_matching, idx_best_matching = 0, None
    for idx_train, training_odour in
        learned_odour_representations.items():
        if best_matching < sum(training_odour==testing_odour):
            best_matching, idx_best_matching =
                sum(training_odour==testing_odour), idx_train
    denoised_testing_data[idx_test] =
        learned_odour_representations[idx_best_matching]
```

**Algorithm 2**: Testing / Denoising (*Python* code). For each testing sample, compare how well the odour representation matches the training samples. Set denoised testing sample as the best-matching training sample.

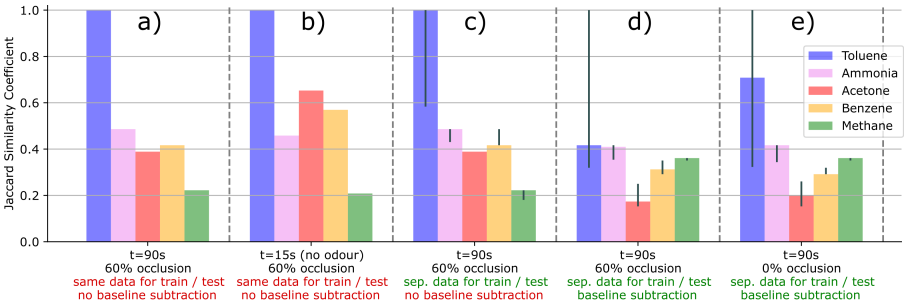


**Figure S1**: Jaccard similarity coefficient of the hash table denoiser's response to occluded Toluene and the learned odour representations. The range of conditions is analogous to the ones described in Fig. 2. Bar heights and error bars respectively display the median and interquartile range of the algorithms' output similarity, as resulting from 72 sensor measurements across 10 Toluene representations. Only five out of 10 gases are depicted for clarity.

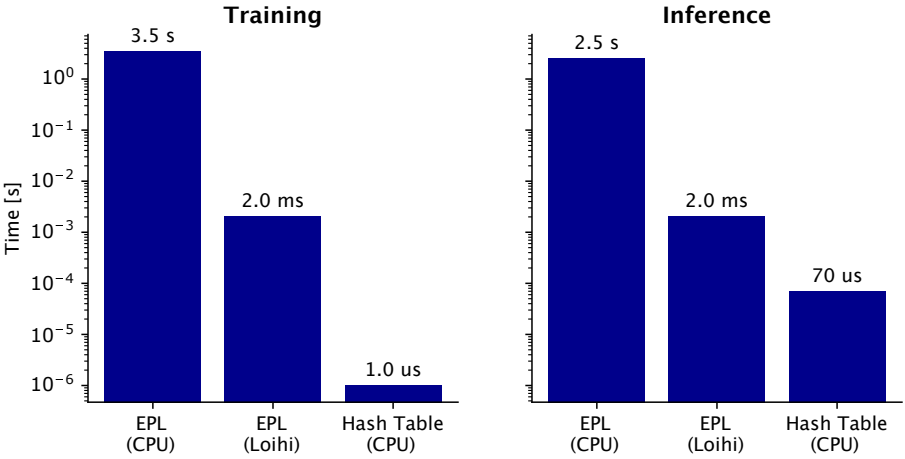


**Figure S2**: Execution time comparison for the different algorithms and platforms. CPU execution times have been computed as the average of 10 and 100 training and inference instances respectively, run on a MacBook Pro M1. The Loihi execution time is denoted in the original paper [1].