NOVEL CONVOLUTION-BASED SIGNAL PROCESSING TECHNIQUE FOR AN ARTIFICIAL OLFATORY MUCOSA

Julian W. Gardner and James E. Taylor
Sensors Research Laboratory, School of Engineering, University of Warwick
Coventry CV4 7AL, UNITED KINGDOM
E-mail: j.w.gardner@warwick.ac.uk

ABSTRACT
Recent advances in our understanding of the biological olfactory system have led to the development of new types of artificial olfactory devices. Inspired by the concept of nasal chromatography (odour components being separated by the olfactory epithelium), the artificial olfactory mucosa or e-mucosa has been conceived. However, the data generated by this device are of a class not encountered in traditional field of artificial olfaction before, due to its spatio-temporal characteristics. A novel processing approach is required to exploit fully the information rich nature of this data. Here a novel convolution based signal processing method is proposed. It is shown how the convolution integral can be applied successfully to different odour data-sets and thus further the development of the artificial or e-mucosa for medical application.

KEY WORDS
Artificial Organ, Biomedical Signal Processing, Artificial Olfactory Mucosa, Convolution Method.

1. Introduction
As our understanding of the biological olfactory system increases, so does our ability to create artificial devices to mimic it. Electronic nose technology is becoming more sophisticated as new discoveries are made in the field of olfaction [1]. The most recent development in the field of artificial olfaction is the development of the artificial olfactory mucosa (AOM) [2], inspired by new understanding of the mechanism and purpose of the nasal chromatography effect [3].

The aqueous layer covering the olfactory epithelium not only protects the olfactory receptors, but also appear to play an important role in the process of olfaction [4]. The separation of odour molecules as they enter and cross the mucous layer creates a temporal difference in their arrival at the olfactory receptors. In addition the partial selectivity of different regions of the olfactory epithelium through different odour receptors creates complex spatio-temporal signals, which are sent to the olfactory bulb and olfactory cortex for odour classification.

The AOM is a device developed to generate a similar spatio-temporal data set, inspired by the biological system (Figure 1). However, this new type of data presents a new challenge to the pattern recognition system. Current approaches to process data are performed using time-independent approaches [5, 6]. Such approaches do not make use of the spatio-temporal nature of the data generated here, and much information useful to classification or quantification is lost. As such, there is a need to develop new, time-dependent approaches, so that this new form of information rich data can be fully exploited.

Figure 1 Schematic representation of the Warwick artificial mucosa and its practical realisation in a portable e-mucosa with hundreds of odour sensors [7,9].

2. The Convolution Method
The convolution processing method is a novel time-independent approach being developed to process artificial mucosa data [7]. This method combines signals from pairs of different odour sensors, which can be separated in space and/or time, and generates a
characteristic signal by application of convolution (Equation 1).

\[ f(t) * g(t) = \int_{-\infty}^{\infty} f(t)g(\tau-t)dt \] (1)

Features are selected and extracted, chosen by the application in which the data are intended to be used (for example, odour classification or quantification/intensity).

Pre-processing can also be used before the convolution processing, if desired and suitable for the intended application.

In the following sections, this convolution approach is applied to several problems: a traditional odour classification problem, a quantification problem and a classification problem for this newly-developed artificial olfactory mucosa device.

### 3. Classification

The first data-set was provided by ISOCS [8], containing data for 4 analytes: anisol, cyclohexanol, propanol and toluene. The data were collected by an array of 8 metal oxide gas sensors (MOX), and an array of 8 quartz microbalance (QMB) coated with different polymer tunings. Each analyte was sampled 40 times.

The data were processed using the convolution method, using auto-scaling pre-processing (Equation 2). The pairings for the convolution included a mixture of intra-array convolutions (MOX * MOX, QMB * QMB) and inter-array convolutions (MOX * QMB).

\[ S^j(t) = \frac{|X^j(t)|}{\max\{X^v_j\}} \] (2)

From the time-dependent convolution signal, several parameters were extracted and used to define time-invariant features, namely: the magnitude of the peak \( H \), the area under the convolution curve \( A \), and the ratio of the area to peak magnitude \( A/H \).

In order to explore classification of the high order data-set, simple principal components analysis (PCA) was used. All processing was carried out in the commercial software MATLAB v7.4.0 (R2007a). Figure 2 shows the PCA plots resulting from this processing. It is evident that the classes are linearly separable in each case because distinct clusters are shown inside the hyper-ellipses. The scores have large values because of the integration over a time scale of many seconds.

An alternative, non-parametric classification approach was also considered, utilising a probabilistic neural network (PNN) to classify the data. The data were split into two sets, one training set and one testing set, comprising of 20 samples each. The network was constructed from the Neural Network Toolbox for commercial software MATLAB v7.4.0 (R2007a).

This classification approach focused on analysing the difference between the various convolution combinations of sensors: MOX only, QMB only, a mixture of intra-array convolutions, inter-array convolutions and a mixture of all possible convolutions. The data were pre-processed using auto-scaling and the same 3 features were selected for classification. Table 1 shows the accuracy of classification of the testing data set.
<table>
<thead>
<tr>
<th></th>
<th>Magnitude</th>
<th>Area</th>
<th>Area/Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOX</td>
<td>96.43%</td>
<td>96.43%</td>
<td>97.62%</td>
</tr>
<tr>
<td>QMB</td>
<td>52.38%</td>
<td>85.71%</td>
<td>97.62%</td>
</tr>
<tr>
<td>Intra-array</td>
<td>97.62%</td>
<td>98.81%</td>
<td>100%</td>
</tr>
<tr>
<td>Inter-array</td>
<td>94.05%</td>
<td>96.43%</td>
<td>100%</td>
</tr>
<tr>
<td>All</td>
<td>96.43%</td>
<td>97.62%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1 Classification accuracy of a probabilistic neural network on test data after training for two different types of sensor arrays. Convolutions were carried out within one type of sensor array (intra) and between the two (inter). The latter has no physical meaning but is interesting to perform.

From these results, it can be seen that the convolution method has successfully classified the data. The strongest feature was the area of the characteristic signal/the peak magnitude of the characteristic signal for both classification approaches. The neural network achieves 100% accuracy, and the strongest approach utilises characteristic signals derived from a mixture of intra-array convolutions, suggesting each sensor type possesses unique and independent information, despite the fundamentally different physical principles of the odour sensors and their dynamical behaviour.

### 4. Quantification

The second data-set provided by the International Society for Olfaction and Chemical Sensing (ISOCS) [8] contained concentration information for the analyte APT at 4 different concentrations, i.e. 28 μg/litre, 30 μg/litre, 33 μg/litre and 36 μg/litre in water solvent. Each concentration was sampled 8 times by an array of 32 polymer sensors, extracted from a 40 ml headspace from 25 ml of solution.

The data were processed using the convolution method; however no pre-processing (i.e. auto-scaling) was performed in order to preserve the concentration information. The sensors were paired for convolution. From the resulting characteristic signals, area/magnitude and magnitude/area were chosen as the two key time-invariant features to be extracted for quantification of the analyte concentration.

Quantification was performed using partial-least squares (PLS) regression, using a non-iterative partial-least squares implemented in MATLAB v7.4.0 (R2007a). The PLS regression was also performed on the raw, unprocessed data for comparison. The results are shown in Figure 3 for the two ratiometric features as well as the raw data, i.e. just the magnitude of the convolution peak.

![Figure 3 PLS plots versus concentration. Each group of 8 samples is represented by a mean, with an error of the standard deviation.](image)

While the group errors are large relative to the regression region in the raw and area/magnitude feature plots, the steeper regression and smaller error regions present in the
magnitude/area shows a distinct improvement in determining the concentration over raw data alone.

5. Artificial Olfactory Mucosa

The AOM used for collection was the so-called Dual-Channel Large Array Electronic Nose that extends the concept to two mucous retention layers [9]. This device utilises 3 polymer sensor arrays, one at the front end, and 2 separated from the front end array by a parallel pair of channels. These channels are coated in a retentive layer, generating a chromatographic effect on the gaseous analytes. One channel is coated with a 5 µm layer of OV-1 (a non-polar stationary phase), and the other channel is coated with a 5 µm layer of Carbowax 20M (a polar stationary phase). This combination of arrays and retentive channels, illustrated in Figure 4, create an information rich spatio-temporal data, and requires a more advanced processing technique.

![Figure 4 Illustration of the layout of the Dual-Channel Large Array Electronic Nose.](image)

Each array consisted of 300 polymer sensors, in 24 different tunings. This creates a data-set that contains much redundancy, and so requires careful handling so that noise effects and co-linearity are kept to a minimum.

Data were collected for 4 essential odours: cinnamon, lemon, lavender and ylang ylang. 5 samples of each analyte were collected. The redundancy of the array allows any clearly incorrect measurements (saturated output responses, or negligible output value), resulting from damaged sensors, can be censored from the data set without compromising the variety of the data. After the censoring process, the remaining data were pre-processed using auto-scaling (Equation 2).

The convolution method was then applied by pairing equivalent sensors between pairings of the arrays: front array × polar array, front array × non-polar array and polar array × non-polar array. From the set of characteristic signal, an area feature was chosen for classification. As before, two classifiers were considered, PCA and PNN, and the processing carried out using commercial software MATLAB v7.4.0 (R2007a).

![Figure 5 PCA plots of the first 3 principal components of AOM data. a) Front Array × Polar Array b) Front Array × Non-Polar Array c) Polar Array × Non-Polar Array.](image)
Figure 5 above shows that PCA is an unsuitable classifier for this data, because there is no clear separation between the samples and so the clusters (indicated by ellipses) are overlapping even in three dimensions.

Figure 6 illustrates the results as pixel images using only the reduced data-set after processing, averaged across one sample. This figure illustrates the spatial nature of the data, and suggests that a non-parametric approach will be much more successful.

Due to the limited number of samples available for use with the PNN, a bootstrap train-and-test approach was used to train the network. A subset of the data is omitted from the data, and the remaining data used to train the network. The accuracy of the network is then evaluated using the omitted subset. This process is then repeated for a significant number of unique subsets to arrive at an overall picture of the performance of the neural net.

However, the censoring process has left an uneven number of sensors across the tunings, with some groups have many sensors remain, and others only a few. Despite this, using all the sensor data is not desirable, as large numbers of redundant, co-linear sensor will only contribute noise to the processing.

In order to select a subset of sensors to use, a Monte Carlo randomiser was employed during the training process. The sensors are randomly arranged, and a single sensor used to train the network. The second sensor is added to the training vector, and the process repeated. If the performance of the classification is improved, the sensor is retained, if not, it is discarded. This is repeated through all valid sensors. This is a sub-optimal, non-exhaustive search method, but has been shown to provide good results in a reasonable timeframe.

The AOM data used here were obtained from a device still at an experimental stage, resulting in a large number of poorly performing sensors. Despite this handicap, the convolution processing has been able to group the data, but the separation between groups is small.

Figure 6 highlights that there are marked differences that a non-parametric classification approach should be able to take advantage of. The F*P accuracy of the PNN classifier was found to be 85%, as shown in Table 2, for this more challenging odour classification problem.

<table>
<thead>
<tr>
<th>Omitted sample</th>
<th>F * NP</th>
<th>F * P</th>
<th>P * NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>50%</td>
<td>100%</td>
<td>75%</td>
</tr>
<tr>
<td>2nd</td>
<td>75%</td>
<td>100%</td>
<td>75%</td>
</tr>
<tr>
<td>3rd</td>
<td>75%</td>
<td>100%</td>
<td>75%</td>
</tr>
<tr>
<td>4th</td>
<td>75%</td>
<td>75%</td>
<td>25%</td>
</tr>
<tr>
<td>5th</td>
<td>100%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Overall:</td>
<td>75%</td>
<td>85%</td>
<td>65%</td>
</tr>
</tbody>
</table>

Table 2 Accuracy of the PNN classification after the bootstrap training process. F is Front sensor array; P is back Polar sensor array, and NP is back Non-polar sensor array.

6. Conclusions

In this work, it has been shown for the first time that the convolution technique can be applied to a variety of problems in the fields of artificial olfaction, including classification, quantification of simple and complex odours. It has also been shown that this processing approach is able to classify data, both in traditional devices in controlled conditions and an experimental artificial olfactory mucosa in less than ideal conditions. A challenging quantification problem shows that correct feature selection can help to improve the performance of this processing method.

Future improvements are needed in the sensor manufacturing to have more reliable and larger sensor signals from the back stages of the e-mucosa. This can then lead to more reproducible sensor responses and these can then be applied to more challenging clinical applications, such as the headspace analysis of urine for bacterial infection.
Acknowledgements

The authors wish to thank Dr. Che Harun for the provision of the raw essential oil data-set and the Engineering and Physical Sciences Research Council (UK) for the financial support of the postgraduate student (James Taylor).

References


