Extracting functional clusters of glomeruli in rat olfactory bulb by non-negative matrix factorization

Takao Yamanaka† and Ricardo Gutierrez-Osuna‡

†Department of Electrical and Electronics Engineering, Sophia University, Tokyo, Japan

‡Department of Computer Science, Texas A&M University, College Station, TX 77843-3112

Correspondence to be sent to:
Takao Yamanaka, Department of Electrical & Electronics Engineering, Sophia University, Japan, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554
email: takao-y@sophia.ac.jp, phone: +81-3-3238-3329, fax: +81-3238-3321

Abstract

Ensemble coding in the early olfactory pathway has been extensively investigated using imaging techniques. These studies have shown that glomeruli with similar affinity gather in close proximity in olfactory bulb, forming a module. In this work, we propose computational methods for analyzing this neural code. Specifically, we show how non-negative matrix factorization (NMF), a machine-learning method for extracting the intrinsic parts of objects, can be used to automatically extract glomerular modules from a database of bulbar activity patterns, as measured with 2-deoxyglucose. The modules extracted by NMF correspond to localized areas in olfactory bulb, in consistency with experimental results from imaging studies on glomerular activity. To validate the emerging representation, we analyzed the relationship between neural activity on these modules and perceptual descriptions of the odorants. We first used pattern-classification techniques to predict ten perceptual descriptors for 53 odorants from their activity on the modules. Our results indicate that NMF is able to extract modules that are intrinsic to the odor coding mechanism. Furthermore, we used mutual information to analyze the relationship between modules and olfactory perception. This analysis revealed the contribution of each module to the olfactory percepts.

Key words

Odor map, neural image analysis, module, 2-deoxyglucose uptake, olfactory perception

Introduction

Olfactory transduction in the rat starts with a large repertoire of odorant receptors (ORs) located on

the cilia of a much larger population of olfactory receptor neurons (ORNs) in the olfactory epithelium [1], [2]. ORNs expressing a given type of OR gene are scattered widely in the epithelium, but their axons exclusively converge onto a few out of a few thousand glomeruli in olfactory bulb [3]-[5], suggesting that an individual glomerulus represents a single type of OR [6]-[8]. As a result of this chemotopic organization, the identity of an odorant is assumed to be represented by a combinatorial activity pattern across glomeruli [9].

These glomerular patterns (i.e., odor maps) have been investigated using imaging techniques, including intrinsic signals [10], [11], activity-related gene expression (c-fos) [12], and glucose uptake measured with radio-labeled 2-deoxyglucose (2-DG) [13], [14]. Analysis of these odor maps for a wide variety of odorants shows that glomeruli with similar affinity are clustered in neighboring regions or modules [15], [16]. Thus, glomerular activity can be described as a pattern across these modules, each of which represents a fundamental functional unit of signal processing in the olfactory system [14], [17]. These studies suggest that much can be learned about olfactory coding by extracting these intrinsic modules from neural activity in the olfactory bulb.

This article presents a computational methodology for extracting and analyzing these modules. The approach is particularly useful when operating with a large number of samples with high dimensions, as is the case with the above imaging studies. At its core, our approach is based on non-negative matrix factorization (NMF), a machine-learning technique that has emerged in recent years as a promising method for analyzing neural and perceptual representations [18]-[20]. NMF has the ability to extract the intrinsic parts of objects because of a non-negative constraint that it imposes on the decomposition (e.g., a face is decomposed into its individual parts: eyes, mouth, nostrils, etc). In the context of olfaction, the intrinsic parts extracted by NMF may correspond to functional modules in the olfactory bulb, where each "functional module" is a local ensemble of neurons that acts as an olfactory coding dimension. Such decomposition cannot be achieved with traditional techniques, such as principal components analysis. Our hypothesis is that, once the fundamental functional units (modules) in the olfactory bulb are extracted, the perceptual properties of a stimulus can be predicted from its neural activity on the olfactory bulb modules. To test this hypothesis, a pattern classifier was used to predict the perceptual properties of the odorants from their activity on the

olfactory-bulb modules extracted by NMF. Among all pattern classification methods (e.g., Bayesian, decision trees, neural networks), Support Vector Machines (SVM) have become the "gold standard" in machine-learning due to their generalization properties, their ability to operate with high-dimensional data, and the absence of local minima in the solution space [21]. Thus, in this work we used SVMs to ensure the best possible prediction results while minimizing the chances that these results could be biased by the limitations of other pattern classifiers. In addition to the pattern-classifier predictions, the statistical relationship between module activities and olfactory perception was analyzed using mutual information [22] in order to visualize the contribution of each module to the olfactory percepts. Unlike the standard (Pearson's) correlation coefficient, which can only measure linear statistical dependencies between two variables, mutual information also measures non-linear dependencies, such as those that are known to exist between neural representations and their percepts.

The contributions of this work are as follow. First, though a number of studies have been reported on the relationship between neural activity and olfactory perception [23]-[27], these studies have concentrated on a small number of odorants. In contrast, our article presents more comprehensive results using a large number of odorants (53) and olfactory perception (10 descriptors). Such comprehensive analysis is complementary to the detailed studies already in the literature, and together can enhance our understanding of the neural mechanisms for odor coding. Second, though principal components analysis (PCA) has been previously used to analyze spatio-temporal patterns in the olfactory bulb [28], [29], this technique is unable to find the intrinsic parts of objects (modules in the olfactory bulb). PCA produces global representations, which are inconsistent with the view that odor coding in the glomerular layer consists of activity in local modules [15], [17]. To the best of our knowledge, ours is the first analysis of olfactory bulb imagery using a non-negative decomposition. Last, we propose a computational methodology that is not limited to a specific animal model, and could be used to automatically extract intrinsic modules when additional data (e.g., in the rat olfactory bulb or the olfactory systems of other phyla) becomes available.

3

Materials and methods

Archive of olfactory bulb images

Images of glomerular activities in olfactory bulb were taken from the Glomerular Activity Response Archive at the University of California at Irvine [30]. The images depict spatial distribution of 2-deoxyglucose uptake evoked in the glomerular layer of rat olfactory bulb in response to a wide range of odorant stimuli. In this method, [^{14}C]2-deoxyglucose is injected into the animal immediately prior to stimulation with an odorized vapor for a duration of up to 45 minutes. The radio-labeled 2-DG is taken into cells through a glucose transporter in proportion to the cells' demand for glucose. Thus, more 2-DG is accumulated in areas that are activated by a stimulus. The 2-DG method has sufficient spatial resolution to detect the activation of a single glomerulus while simultaneously recording activity across an entire glomerular layer [30], [31].

In this paper, all the images available in the archive as of March 2005 were used for the analysis. These include 172 images of 121 odorants with replicated measurements and multiple concentrations for some odorants. The representatives for the odorants used in this study are listed in Table 1. In order to reduce computations, each image was subsampled from its original size (approx. 220×400 pixels) down to 73×132 pixels, and cropped to preserve the region of interest common to all images. Subsequently, each image was normalized by its L2 norm in order to remove concentration effects, and each pixel was normalized by its variance across all odorants to remove biases towards stronger glomeruli. In this paper, only dorsal-centered images were employed.

⟶ **[Table 1]**

Module extraction

Olfactory bulb modules were extracted using the NMF algorithm proposed by Lee and Seung [19], [20]. NMF is a bilinear decomposition with a non-negativity constraint on the elements of the decomposition matrices:

$$V_{ij} \approx \left(\mathbf{WH}\right)_{ij} = \sum W_{ia} H_{aj} \qquad\qquad W_{ij} \geq 0, H_{ij} \geq 0 \qquad\qquad (1)$$

where $\mathbf{V}$ is a $n \times m$ data matrix, each column being an n-dimensional sample out of a dataset with m samples. For image datasets, each column of $\mathbf{V}$ is obtained by sampling the corresponding image in raster-scan fashion (i.e., by concatenating columns in the image into a single column). The r columns of $\mathbf{W}$ are called basis vectors; since each represents a basis of the vector space used to represent the original data. In turn, each column of $\mathbf{H}$ is called an encoding vector, and is in one-to-one correspondence with a sample vector in $\mathbf{V}$: each encoding vector in $\mathbf{H}$ represents the coefficients of a linear combination of the basis vectors in $\mathbf{W}$ for a particular sample vector in $\mathbf{V}$. In NMF, the basis vectors $\mathbf{W}$ and the encoding vectors $\mathbf{H}$ are obtained from the data matrix $\mathbf{V}$ by minimizing the reconstruction error $|\mathbf{V} - \mathbf{WH}|$ with a non-negativity constraint on each element of $\mathbf{W}$ and $\mathbf{H}$. Since no subtractions can occur in $\mathbf{WH}$, NMF operates according to the intuitive notion of combining parts to form a whole. Thus, NMF has the ability to learn basis vectors that are intrinsic parts of objects; e.g., when applied to facial images, NMF returns local regions such as mouth, eyes, nostrils, etc. For our purposes, each basis vector in $\mathbf{W}$ corresponds to a glomerular module, whereas each encoding vector in $\mathbf{H}$ corresponds to the activity of all modules for a particular odorant (i.e., a pattern or a code).

The original NMF algorithm of Lee and Seung [20] provides an efficient iterative procedure for the calculation of $\mathbf{W}$ and $\mathbf{H}$ from random initialization, but the solutions are rather sensitive to the choice of initial values. To address this issue, our implementation of the NMF algorithm uses a deterministic initialization driven by the distribution of the data. First, a k-means clustering algorithm combined with an initialization method based on principal components analysis [32] is used to identify representative clusters of pixels based on their response across all odorants. The centers of the resulting clusters are used as initial encoding vectors $\mathbf{H}$ for NMF. The initial basis vectors $\mathbf{W}$ for NMF are then obtained by minimizing the reconstruction error $|\mathbf{V} - \mathbf{WH}|$ for the initial encoding vectors $\mathbf{H}$ using a non-negative least squares algorithm [33]. The final NMF decomposition is obtained from these initial matrices using the original NMF.

Relationship between module activity and olfactory perception

To validate the extracted modules, we investigated the extent to which neural activity in the rat olfactory-bulb modules can be used to predict perceptual properties of the odorants, obtained from Flavornet [34]. Although there is no evidence that the olfactory perception spaces of rats and humans are the same, recent genetic studies [2] indicate that the rat OR repertoire covers a similar "receptor space" as the human OR repertoire.

The Flavornet dataset is relatively sparse, with only a few descriptors (typically 1-3) available for each odorant and vice versa. Among all the Flavornet descriptors, our analysis was conducted only on those which were active for more than five bulb images in the archive. Table 2 shows the ten descriptors and the odorant indices (from Table 1) for which they are active; replicates of an odorant index indicate multiple concentrations or measurements. Although NMF modules were computed on all 121 odorants (172 images) in the archive, only 53 odorants (87 images) were used to correlate neural and perceptual representations. The 53 odorants are those for which at least one of their active descriptors is also active for another odorant(s); this restriction filters out odorants whose descriptors may be too specific.

⟶ **[Table 2]**

First, the relationship between module activities and odor properties was examined as a classification paradigm. Each olfactory bulb image (m=87) was assigned an indicator variable for each odor descriptor: "+1" if the descriptor was active for the odorant, and "-1" otherwise. The objective was to predict the value of each indicator variable based on the module activities. For this purpose, a support vector machine (SVM) classifier was used [35], [36]. Given that classification performance of the original SVM algorithm deteriorates when there is a large mismatch in the numbers of data for each group (imbalanced dataset), we employed an SVM algorithm that allows different error costs to be defined for each class [37]. Following [37], the ratio between the error cost for the positive and negative classes was set to the inverse of the imbalance ratio (the ratio between the number of data in the positive and negative classes). A radial basis function was employed as the SVM kernel function.

Two metrics, *sensitivity* and *specificity*, were used for evaluating the prediction results. The *sensitivity* is the true positive rate, defined as the proportion of positive instances that are correctly classified as positive. The *specificity* is the true negative rate, defined as the proportion of negative instances that are correctly classified as negative. These two metrics are combined into a geometric mean (*g-mean*), defined as the square root of the product of sensitivity and specificity [37], [38]. Sensitivity, specificity, and g-mean were calculated using 10-fold cross validation [35]; for each fold, 90% of the data was used to train the SVM classifiers, and the remaining 10% to estimate the three figures of merit.

In order to test the hypothesis that the perceptual properties of odorants can be predicted from module activity, the performance of the NMF modules was compared against a baseline "chance level". To obtain chance levels, the indicator value (±1) for each olfactory descriptor was randomly assigned for all odorants, while ensuring that the number of (randomly) active odorants for each descriptor was identical to that in the original dataset. An SVM classifier was then built to predict the value of each (random) descriptor from the NMF module activities. The final chance level was obtained as the average across ten separate experiments, each with a different random assignment of descriptors. The performance on the NMF modules was evaluated against that of two additional SVM classifiers: one that predicted odor percepts from the original modules in the archive, shown in Figure 6(b), and a second one that used "local" modules determined based on spatial locality (i.e., no chemotopic considerations). The latter, shown in Figure 6(c), were obtained by k-means clustering of pixels according to their position in the image.

Second, we also employed mutual information [22] to visualize the relationship between module activities and odor descriptors. Mutual information is a measure of statistical dependence between two random variables based on the concept of entropy, a measure of uncertainty or disorder in a random variable. In the context of this paper, the two random variables are the activity of each module and the indicator variable (±1) for each descriptor. For a discrete random variable *X*, the entropy *H*(*X*) is defined as:

$$H(X) = -\sum_{x} p(x) \log p(x) \qquad\qquad H(X) \geq 0 \qquad\qquad (2)$$

where *p*(*x*) is the probability mass function at *X*=*x*. In turn, the mutual information between two random

variables *X* and *Y* is defined as:

$$
\begin{aligned}
I(X;Y) &= \sum_x \sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= H(X) - H(X/Y) \qquad\qquad I(X;Y) \geq 0 \qquad\qquad (3) \\
&= H(Y) - H(Y/X)
\end{aligned}
$$

where $p(x, y)$ is the joint probability, and $H(X|Y)$ is the conditional entropy, the remaining uncertainty in *X* given that *Y* is known:

$$
H(X/Y) = -\sum_x \sum_y p(x,y)\log p(x/y) \qquad H(X/Y) \geq 0 \qquad\qquad (4)
$$

The mutual information $I(X; Y)$ can be interpreted as the reduction in the uncertainty of *X* when the random variable *Y* is determined (or vice versa, due to the symmetry on *X* and *Y*). Thus, mutual information increases with the statistical dependence between *X* and *Y*. Moreover, since the mutual information is upper-bounded by the entropies of *X* and *Y* (refer to Eq. 3), in this work it was normalized by the lower of the two entropies as:

$$
NMI(X;Y) = \frac{I(X;Y)}{\min(H(X),H(Y))} \qquad\qquad (5)
$$

Thus, $NMI(X;Y)$ is bounded [0, 1] for ease of interpretation. In order to compute the mutual information, the probabilities $p(x)$, $p(y)$, and $p(x, y)$ were estimated as histograms, with quantization levels of 2 for each descriptor (± 1) and 20 for the activity of each module.

Results

Module extraction

Examples of olfactory bulb images from the archive [30] are shown in Figure 1, where the first rows in (a), (b), and (c) show the neural representation for odorants of fruit, fat, and sweat smell, respectively. Lighter parts in the images represent areas of high neural activity, as detected by the 2-DG method. As shown in the images, the olfactory bulb displays similar activity distribution for odorants with the same smell

impression. The 172 olfactory bulb images in the archive were used to extract basis vectors **W** using NMF. The activity and contours of the resulting modules are shown in Figures 2 (a) and 6 (a), respectively. The number of the basis vectors was set to 16, the same as the number of experimentally-extracted modules in the archive, so that an unbiased comparison between NMF modules and archive modules could be performed. For comparison purposes, Figure 2 (b) shows the basis vectors extracted with principal components analysis (PCA), the classical representative among all dimensionality-reduction techniques. PCA minimizes the reconstruction error in the least-squares sense which, in general, requires negative values on elements of the basis vectors [39].

$\longrightarrow$ **[Figure 1, Figure 2]**

NMF operates in an unsupervised fashion. Illustrated in Figure 3, NMF finds a decomposition of the data into a series of basis images (or modules) and encoding vectors (or module activities), neither of which is allowed to be negative. Since subtractions are not allowed in the decomposition, basis images tend to be sparse, that is, many entries of the basis matrix are close to zero. This sparseness of the basis matrix leads to a local representation for the modules. PCA also operates in an unsupervised fashion. However, PCA allows the modules (eigenvectors) and encodings (principal components) to be negative, which leads to a global representation of the modules, as shown in Figure 2 (b).

$\longrightarrow$ **[Figure 3]**

The encoding vectors **H** (or module activities) produced by NMF can be thought of as the coefficients of a linear superposition in terms of basis vectors **W**. Module activities for images with fruit, fat, and sweat smell descriptors are shown in Figures 4 (a), (b), and (c), respectively. The module indices in Figure 4 correspond to those in Figure 2 (a). The fruit smell caused high activation primarily in the 10th module, whereas the fat smell showed high activation in the 6th and/or 14th modules, and the sweat smell produced high activity in the 1st and 3rd modules. Thus, smell impressions can be tied to areas of high neural activity in olfactory bulb.

$\longrightarrow$ **[Figure 4]**

Reconstructions of the original images from the NMF decomposition are shown in the lower rows of

Figure 1. These results indicate that, in addition to extracting the local parts, NMF is also able to represent the original images with a relatively low number of basis vectors by removing high frequency noise.

Relationship between module activities and olfactory perception

Based on the module activities obtained by NMF, the smell impression was predicted using an SVM classifier. Prediction results for 10 fold-cross validation are shown in Figure 5, in terms of the sensitivity (true positive rate), the specificity (true negative rate), and the g-mean (combined measure of sensitivity and specificity). As shown in Figure 5 (a), the predictability for the NMF modules was much higher than the chance level, supporting the hypothesis that olfactory percepts can be predicted from neural activity in olfactory bulb.

In order to analyze the effectiveness of the NMF method, the prediction result from NMF modules (Figure 6a) was also compared with those for archive modules (Figure 6b) and local modules obtained by "mosaicing" the bulb spatially (Figure 6c).   The three sets of modules were compared by their ability to predict each of 10 olfactory descriptors. As discussed earlier, a support vector machine (SVM) classifier was built to predict each olfactory descriptor (10 descriptors) from each set of modules (3 sets), for a total of 30 classifiers. Figure 5 (b) shows the accuracy of these predictions on test samples in terms of (1) sensitivity (true positive rate), (2) specificity (true negative rate), and the geometric mean of these two. The average g-mean across the 10 descriptors was 81.1% for NMF modules, 74.9% for archive modules, and 76.3% for local modules.

To determine whether or not differences in accuracy between NMF modules and archive modules were statistically significant, the g-mean of each representation across the ten descriptors was treated as a population, and these two populations (ten samples each) were compared using a paired t-test. Similar comparisons were performed between NMF and local modules, and between archive and local modules The paired t-test reveals significant differences at the 0.05 level between NMF and archive modules ($p = 0.0455$, $t = 2.32$), and between NMF and local modules ($p = 0.0440$, $t = 2.34$). Moreover, differences between the archive modules and the local modules were found *not* to be statistically different ($p = 0.5441$, $t = 0.63$).

A similar procedure could have been used to compare the performance across different descriptors. This comparison would have allowed us to determine whether or not some olfactory descriptors are easier to predict than others. However, each of the ten descriptors is described by a population of only three samples (i.e., three sets of modules), which was deemed to be insufficient. Therefore, no comparisons were performed among descriptors.

⟶ **[Figure 5, Figure 6]**

Furthermore, the relationship between module activities and odor descriptors was visualized by computing the normalized mutual information, which can be thought of as a non-linear measure of correlation. Results are shown in Figure 7, where the size of each circle is proportional to the normalized mutual information. These results are consistent with those in Figure 4, which instead showed the activity of NMF modules. There are dependencies between fruit smells and activity in module 10, between fat smells and activity in modules 6/14, and between sweat smell and activity in modules 1 and 3. Furthermore, other relationships can be also observed in Figure 7, indicating that the normalized mutual information is a useful tool to analyze the neural coding of smell impressions in olfactory bulb.

⟶ **[Figure 7]**

Discussion

The following conclusions can be extracted from the experimental results obtained in the previous section. First, the perceptual properties of an odor (as indicated by the ten olfactory descriptors) can be predicted from neural activity in the glomerular layer. This conclusion is clearly illustrated in Figure 5 (a), which shows that the prediction performance of olfactory percepts from neural activity is much higher than chance level. Second, NMF is an effective method for extracting intrinsic modules, this is, local representations that are highly predictive of the olfactory percepts. Supporting this conclusion is the experimental comparison between NMF modules (Figure 2a, Figure 6a) and three additional representations: PCA modules (Figure 2b), archive modules (Figure 6b), and a naïve partition into purely "spatial" modules (Figure 6c). As discussed earlier, the global representation of PCA is inconsistent with the view that odor

11

coding in the glomerular layer consists of activity in local modules [15], [17]. Among the three local representations (i.e., NMF, archive, "spatial" modules), NMF provides the highest prediction results, as shown in Figure 5 (b). Third, mutual information provides a high-level view of the relationship between modules and olfactory perception (i.e., the ensemble code). As shown in Figure 7, module #1 appears to be associated with a very specific percept ("sweat"), whereas module #8 is associated with several, broader percepts ("fruit", "fat", "sweet"). Likewise, the results in Figure 7 suggest that broader percepts may be associated with widespread activity in olfactory bulb, though further studies are required to test this hypothesis.

Comparison between NMF and archive modules

Most of the modules extracted by NMF (Figure 2a, Figure 6a) show two clusters in lateral and medial parts of olfactory bulb. This is consistent with the modules provided with the archive, shown in Figure 6 (b), where each module was labeled by a pair of upper-case and lower-case letters [13], [30]. Considering that the NMF modules also predict the perceptual properties of the odorants more accurately than the archive modules, we can conclude that NMF provides an objective method for extracting intrinsic modules in an automated fashion from a large number of samples.

For comparison purposes, the number of NMF modules used in this article was set to the same value as that in the archive (16 modules). However, given that the number of archive modules was determined empirically, it is possible that the actual number of intrinsic modules may be larger than the one used here. Recent studies have estimated the OR repertoire to be around 1,000 in the mouse, and 500-750 in humans [2]. Thus, these numbers serve as an upper bound on the number of intrinsic modules in olfactory bulb. However, given that glomeruli with similar affinity gather in close proximity [15], [16], the number of intrinsic modules is likely to be lower than the OR repertoire. Determining the number of intrinsic modules in the rat olfactory bulb is an important topic for future research. Once additional olfactory images from a sufficient number of odorants are available, the proposed method can be used to generate estimates of the number and location of these functional modules.

Relationship between module activities and olfactory perception

The result in Figure 5 (a) supports the hypothesis that olfactory perception can be predicted from neural activity in olfactory bulb, whereas the result in Figure 5 (b) demonstrates the effectiveness of the NMF method in extracting the intrinsic signal-processing units of olfaction. Figure 5 (b) also yields an interesting result: the "local" non-chemotopic modules in Figure 6 (c) have similar predictive accuracy as the archive modules in Figure 6 (b). One possible explanation for this result is that the olfactory bulb has a larger number of intrinsic modules, each composed of closely located glomeruli. If such were the case, the three sets of modules considered in this article would only provide a very coarse representation, one that averages away information in highly localized modules. This issue deserves further investigation, which could be greatly facilitated by the computational methods in this article.

Several studies have been reported on the relationship between neural activity and olfactory perception [13]. If odorants are coded using unique patterns of glomerular activity, it should then be possible to predict smell percepts from these activity patterns. To test this hypothesis, Linster et al. [23] and Rubin and Katz [24] used pairs of enantiomers (optical isomers). Their behavioral experiments revealed that rats can discriminate between enantiomers only if their glomerular activities are different. The relationship between neural activity and olfactory perception has also been previously investigated using mucosal activities in rats [25], [26] and antennal lobe activities in honeybees [27] using 5-16 odorants. Compared with these studies, our article presents results on a large number of odorants (53) and olfactory perception (10 descriptors). Therefore, the results in this paper provide a comprehensive insight of the relationship between neural and perceptual spaces. Such comprehensive analysis is complementary to the detailed studies mentioned above, and together can enhance our understanding of the neural mechanisms for odor coding.

The perceptual descriptors used in this paper were obtained from the Flavornet database [34], which contains the smell impression of humans. Though it is not clear whether the olfactory perception space of humans and rats are similar, our study is supported by a number of comparative studies. Olfactory systems across most species in most phyla show striking neuroanatomical and molecular similarities, which support

13

the idea of a common underlying design [40], [41]. As mentioned earlier, cluster analysis of human and rat OR genes shows that the rat OR repertoire cover a similar "receptor space" as the human OR repertoire [2]. Thus, the high predictive accuracy of our SVM classifier (summarized in Figure 5) is supportive of the hypothesis that humans and rats may also share a similar perceptual space.

This article has analyzed the relationship between neural activities and olfactory perception; an important direction of future work is the joint analysis of neural activity and molecular features of the odorant. Recent work by Leon et al. [13], [42], [43] revealed that the molecular features most relevant to neural activities are functional groups, hydrocarbon structure, and functional group position on a molecule. The computational methods presented in this paper may prove beneficial to investigating additional relevant molecular features, and may reveal a larger number of modules than those found here. Such investigations could provide key insights into the thus-far elusive structure-odor relationships [44], [45].

Acknowledgments

References

[1] L. Buck and R. Axel, "A novel multigene family may encode odorant receptors: a molecular basis for odor recognition," *Cell*, vol. 65, pp. 175-187, 1991.

[2] X. Zhang and S. Firestein, "The olfactory receptor gene superfamily of the mouse," *Nat. Neurosci.*, vol. 5, pp. 124-133, 2002.

[3] K. J. Ressler, S. L. Sullivan, and L. B. Buck, "Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb," *Cell*, vol. 79, pp. 1245-1255, 1994.

[4] R. Vassar, S. K. Chao, R. Sitcheran, J. M. Nunez, L. B. Vosshall, and R. Axel, "Topographic organization of sensory projections to the olfactory bulb," *Cell*, vol. 79, pp. 981-991, 1994.

[5] P. Mombaerts, F. Wang, C. Dulac, S. K. Chao, A. Nemes, M. Mendelsohn, J. Edmondson, and R. Axel, "Visualizing an olfactory sensory map," *Cell*, vol. 87, pp. 675-686, 1996.

[6] L. B. Buck, "The molecular architecture of odor and pheromone sensing in mammals," *Cell*, vol. 100, pp. 611-618, 2000.

[7] P. Mombaerts, "Seven-transmembrane proteins as odorant and chemosensory receptors," *Science*, vol. 286, pp. 707-711, 1999.

[8] K. Mori, H. Nagao, and Y. Yoshihara, "The olfactory bulb: coding and processing of odor molecule information," *Science*, vol. 286, pp. 711-715, 1999.

[9] K. Mori, "Grouping of odorant receptors: odour maps in the mammalian olfactory bulb," *Biochem. Soc. Trans.*, vol. 31, pp. 134-136, 2003.

[10] N. Uchida, Y. K. Takahashi, M. Tanifuji, and K. Mori, "Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features," *Nat. Neurosci.*, vol. 3, pp. 1035-1043, 2000.

[11] M. Meister and T. Bonhoeffer, "Tuning and Topography in an odor map on the rat olfactory bulb," *J. Neurosci.*, vol. 21, pp. 1351-1360, 2001.

[12] K. M. Guthrie, A. J. Anderson, M. Leon, and C. Gall, "Odor-induced increases in c-fos mRNA expression reveal an anatomical "unit" for odor processing in olfactory bulb," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, pp. 3329-3333, 1993.

[13] M. Leon and B. A. Johnson, "Olfactory coding in the mammalian olfactory bulb," *Brain Res. Rev.*, vol. 42, pp. 23-32, 2003.

[14] B. A. Johnson, H. Farahbod, Z. Xu, S. Saber, and M. Leon, "Local and global chemotopic organization: general features of the glomerular representations of aliphatic odorants differing in carbon number," *J. Comp. Neurol.*, vol. 480, pp. 234-249, 2004.

[15] Y. K. Takahashi, M. Kurosaki, S. Hirono, and K. Mori, "Topographic representation of odorant molecular features in the rat olfactory bulb," *J. Neurophysiol.*, vol. 92, pp. 2413-2427, 2004.

[16] K. M. Igarashi and K. Mori, "Spatial representation of hydrocarbon odorants in the ventrolateral zones of the rat olfactory bulb," *J. Neurophysiol.*, vol. 93, pp. 1007-1019, 2005.

[17] B. A. Johnson and M. Leon, "Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration," *J. Comp. Neurol.*, vol. 422, pp. 496-509, 2000.

[18] P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111-126, 1994.

[19] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, 1999.

[20] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2001.

[21] N. Cristianini and J. Shawe-Taylor, *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[22] G. Deco and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*. Springer, 1995.

[23] C. Linster, B. A. Johnson, E. Yue, A. Morse, Z. Xu, E. E. Hingco, Y. Choi, M. Choi, A. Messiha, and M. Leon, "Perceptual correlates of neural representations evoked by odorant enantiomers," *J. Neurosci.*, vol. 21, pp. 9837-9843, 2001.

[24] B. D. Rubin and L. C. Katz, "Spatial coding of enantiomers in the rat olfactory bulb," *Nat. Neurosci.*, vol. 4, pp. 355-356, 2001.

[25] P. F. Kent, S. L. Youngentob, and P. R. Sheehe, "Odorant-specific spatial patterns in mucosal activity predict perceptual differences among odorants," *J. Neurophysiol.*, vol. 74, pp. 1777-1781, 1995.

[26] P. F. Kent, M. M. Mozell, S. L. Youngentob, and P. Yurco, "Mucosal activity patterns as a basis for olfactory discrimination: comparing behavior and optical recordings," *Brain Res.*, vol. 981, pp. 1-11, 2003.

[27] F. Guerrieri, M. Schubert, J. C. Sandoz, and M. Giurfa, "Perceptual and neural olfactory similarity in honeybees," *PLoS. Biol.*, vol. 3, pp. 718-732, 2005.

[28] R. Friedrich and G. Laurent, "Dynamic optimization of odor representation by slow temporal patterning of mitral cell activity," *Science*, vol. 291, pp. 889-894, 2001.

[29] R. F. Galan, S. Sachse, C. G. Galizia, and A. V. M. Herz, "Odor-driven attractor dynamics in the antennal lobe allow for simple and rapid olfactory pattern classification," *Neural Comput.*, vol. 16, pp. 999-1012, 2004.

[30] M. Leon and B. A. Johnson, *Glomerular activity response archive*. http://leonserver.bio.uci.edu/, 2005.

[31] B. A. Johnson, C. C. Woo, and M. Leon, "Spatial coding of odorant features in the glomerular layer of the rat olfactory bulb," *J. Comp. Neurol.*, vol. 393, pp. 457-471, 1998.

[32] T. Su and J. Dy, "A deterministic method for initializing k-means clustering," *16th IEEE International Conference on Tools with Artificial Intelligence*, pp. 784-786, 2004.

[33] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Prentice-Hall, Chapter 23, p. 161, 1974.

[34] T. Acree and H. Arn, *Flavornet*. http://www.flavornet.org/, 2004.

[35] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[36] S. R. Gunn, *MATLAB Support Vector Machine Toolbox*. http://www.isis.ecs.soton.ac.uk/isystems/kernel/, 1998.

[37] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," *European Conference on Machine Learning*, 2004.

[38] M. Kubat, R. Holte, and S. Matwin, "Learning when negative examples abound," *European Conference on Machine Learning*, 1997.

[39] R. O. Duda, P. E. Hart, and D. G Stork, *Pattern Classification*. A Wiley-Interscience Publication, 2001.

[40] J. G. Hildebrand and G. M. Shepherd, "Mechanisms of olfactory discrimination: converging evidence for common principles across phyla," *Annu. Rev. Neurosci.*, vol. 20, pp. 595–631, 1997.

[41] N. J. Strausfeld and J. G. Hildebrand, "Olfactory systems: common design, uncommon origins?" *Curr. Opin. Neurobiol.*, vol. 9, pp. 634-639, 1999.

[42] B. A. Johnson, H. Farahbod, and M. Leon, "Interactions between odorant functional group and hydrocarbon structure influence activity in glomerular response modules in the rat olfactory bulb," *J. Comp. Neurol.*, vol. 483, pp. 205-216, 2005.

[43] B. A. Johnson, H. Farahbod, S. Saber, and M. Leon, "Effects of functional group position on spatial representations of aliphatic odorants in the rat olfactory bulb," *J. Comp. Neurol.*, vol. 483, pp. 192-204, 2005.

[44] S. S. Schiffman, "Physicochemical correlates of olfactory quality," *Science*, vol. 185, pp. 112-117, 1974.

[45] K. J. Rossiter, "Structure-Odor Relationships," *Chem. Rev.*, vol. 96, pp. 3201-3240, 1996.

Table Captions

Table 1 List of representatives for odorants.

Table 2 Smell descriptors used for the analysis, selected from Flavornet [34]. Only descriptors which have more than 5 images in the olfactory bulb archive were employed.

Figure Captions

Fig. 1. Olfactory bulb images for odorants with similar smell impreassion. (a) 1-Fruit, (b) 3-Fat, (c) 6-Sweat. The first row in each figure shows original olfactory bulb images. The second row shows the images reconstructed based on the extracted NMF modules and their activities. Odorant numbers are shown in Table 1. For multiple images of an odorant, only a representative image is shown. The maximum and minimum values in each image are represented by white and black, respectively.

Fig. 2. Extracted modules by (a) NMF and (b) PCA. For PCA, the first 16 eigenvectors of the covariance matrix of the data matrix are presented. In (a), dashed line represents the contour of olfactory bulb. The maximum and minimum values in each image (absolute values for PCA) are represented by white and black, respectively.

Fig. 3. Illustration of the decomposition performed by non-negative matrix factorization. Activity in the olfactory bulb (left) is represented as the sum of activities (right; shown as a 4x4 matrix, where lighter colors denote higher values) across sixteen modules (center; shown as a 4x4 image mosaic).

Fig. 4. NMF module activities to odorants with similar smell impression. (a) 1-Fruit, (b) 3-Fat, (c) 6-Sweat. Module numbers correspond to the numbers in Figure 2 (a). For multiple images of an odorant, only the activity of a representative image is shown.

Fig. 5. Prediction of smell impression from NMF module activities using a Support Vector Machine classifier. (a) Comparison of NMF with chance level. "Random" represents a chance level: prediction results from NMF module activity when descriptors are randomly assigned to images (The number of active images is kept the same as the original for each descriptor). (b) Comparison of NMF with archive modules (Figure 6b) and local modules (Figure 6c). "Local" modules were determined by dividing the olfactory bulb region only based on spatial location, as shown in Figure 6 (c). Top: Sensitivity (true positive rate), middle: Specificity (true negative rate), bottom: g-mean . $\left(\sqrt{\text{Sensitivity} \times \text{Specificity}}\right)$ The difference between "NMF" and "Archive", and the difference between "NMF" and "Local" were statistically significant at the 0.05 level in paired t-tests.

Fig. 6. (a) Outline of NMF modules in Figure 2 (a). (b) Modules in the Glomerular Activity Response Archive [30]. (c) Local modules determined by dividing the olfactory bulb region based on spatial location.

Fig. 7. Relationship between descriptors and NMF module activities, visualized by normalized mutual information.

Table 1 List of representatives for odorants.

#	Odorant	CAS #	#	Odorant	CAS #
1	1-heptanol	111-70-6	64	ethyl caproate	123-66-0
2	1-hexanol	111-27-3	65	ethyl heptanoate	106-30-9
3	1-nonanol	143-08-8	66	ethyl octanoate	106-32-1
4	1-octanol	111-87-5	67	ethyl propionate	105-37-3
6	1-propanol	71-23-8	68	ethyl valerate	539-82-2
11	2-heptanone	110-43-0	72	guaiacol	90-05-1
13	2-hexanone	591-78-6	73	heptanal	111-71-7
17	2-nonanone	821-55-6	74	hexanal	66-25-1
18	2-octanol	123-96-6	75	hexyl acetate	142-92-7
19	2-octanone	111-13-7	80	isovaleric acid	503-74-2
20	2-pentanol	6032-29-7	91	methyl benzoate	93-58-3
21	2-pentanone	107-87-9	92	methyl caproate	106-70-7
22	2-undecanone	112-12-9	94	methyl cyclohexanecarboxylate	4630-82-4
25	3-hexanone	589-38-8	102	nonanal	124-19-6
33	4-methyl-3-penten-2-one	141-79-7	104	octanal	124-13-0
40	acetophenone	98-86-2	106	octanoic acid	124-07-2
45	benzaldehyde	100-52-7	107	p-anisaldehyde	123-11-5
46	benzyl alcohol	100-51-6	108	pentanal	110-62-3
50	butyric acid	107-92-6	109	propionic acid	79-09-4
51	caproic acid	142-62-1	120	valeric acid	109-52-4
61	decanal	112-31-2			

Table 2 Smell descriptors used for the analysis, selected from Flavornet [34]. Only descriptors which have more than 5 images in the olfactory bulb archive were employed.

	Descriptor	# of odorants	# of images	Odorant # [a]
1	Fruit	9	9	21, 64, 65, 66, 67, 68, 75, 92, 94
2	Green	8	11	1, 1, 1, 2, 2, 3, 17, 20, 22, 102, 104
3	Fat	7	8	3, 18, 66, 73, 73, 74, 102, 104
4	Sweet	6	6	33, 46, 72, 91, 92, 107
5	Soap	5	11	11, 11, 17, 19, 19, 19, 19, 19, 61, 61, 104
6	Sweat	5	14	50, 51, 51, 51, 80, 106, 120, 120, 120, 120, 120, 120, 120, 120
7	Chemical	3	6	1, 1, 1, 4, 4, 33
8	Pungent	3	7	6, 108, 108, 108, 108, 108, 109
9	Ether	3	8	13, 13, 13, 13, 13, 13, 21, 25
10	Almond	3	8	40, 45, 45, 108, 108, 108, 108, 108

[a] Duplicate numbers represent multiple images for the odorant with different concentrations or measurements.

Fig. 1. Olfactory bulb images for odorants with similar smell impression. (a) 1-Fruit, (b) 3-Fat, (c) 6-Sweat. The first row in each figure shows original olfactory bulb images. The second row shows the images reconstructed based on the extracted NMF modules and their activities. Odorant numbers are shown in Table 1. For multiple images of an odorant, only a representative image is shown. The maximum and minimum values in each image are represented by white and black, respectively.

(a)   (b)

Fig. 2. Extracted modules by (a) NMF and (b) PCA. For PCA, the first 16 eigenvectors of the covariance matrix of the data matrix are presented. In (a), dashed line represents the contour of olfactory bulb. The maximum and minimum values in each image (absolute values for PCA) are represented by white and black, respectively.

ORIGINAL
IMAGE

BASIS IMAGES
(MODULES)

ENCODING VECTOR
(MODULE ACTIVITY)

Fig. 3. Illustration of the decomposition performed by non-negative matrix factorization. Activity in the olfactory bulb (left) is represented as the sum of activities (right; shown as a 4x4 matrix, where lighter colors denote higher values) across sixteen modules (center; shown as a 4x4 image mosaic).

Fig. 4. NMF module activities to odorants with similar smell impression. (a) 1-Fruit, (b) 3-Fat, (c) 6-Sweat. Module numbers correspond to the numbers in Figure 2 (a). For multiple images of an odorant, only the activity of a representative image is shown.

Fig. 5. Prediction of smell impression from NMF module activities using a Support Vector Machine classifier. (a) Comparison of NMF with chance level. "Random" represents a chance level: prediction results from NMF module activity when descriptors are randomly assigned to images (The number of active images is kept the same as the original for each descriptor). (b) Comparison of NMF with archive modules (Figure 6b) and local modules (Figure 6c). "Local" modules were determined by dividing the olfactory bulb region only based on spatial location, as shown in Figure 6 (c). Top: Sensitivity (true positive rate), middle: Specificity (true negative rate), bottom: g-mean $\left(\sqrt{\text{Sensitivity} \times \text{Specificity}}\right)$. The difference between "NMF" and "Archive", and the difference between "NMF" and "Local" were statistically significant at the 0.05 level in paired t-tests.

Fig. 6. (a) Outline of NMF modules in Figure 2 (a). (b) Modules in the Glomerular Activity Response Archive [30]. (c) Local modules determined by dividing the olfactory bulb region based on spatial location.

Fig. 7. Relationship between descriptors and NMF module activities, visualized by normalized mutual information.