Analysis of an Insect's Olfactory Receptor Neuron Response by NMF Method for Odor Approximation

Yuki Harada and Takamichi Nakamoto*

Tokyo Institute of Technology,
Nagatsuta 4259, Midori-ku, Yokohama, Kanagawa 226-8503, Japan

Odor approximation to express a variety of odors using a small number of odor components is important. Our group has for years been studying odor approximation to express a variety of odors using a small number of odor components. Basis vectors corresponding to odor components can be extracted from an established database by the nonnegative matrix factorization (NMF) method. We have built a database of several hundred types of essential oil using mass spectrometry (MS). We have already reported that the approximated odors of essential oils can be reproduced by blending odor components. However, it is expected that the approximation in biological space provides higher accuracy. It is necessary to compare mass-spectrum data space with the biological space. In this study, we also used the NMF method to analyze the biological sensing space on the basis of insects' response data. Odor chemicals should be located spatially as characterized by 24 olfactory receptor neuron (ORN) responses. When chemical structures are much different, the distance between them is long. We found that the extracted basis vectors were located according to the spatial distribution of odors. Analysis of the relationship between the number of basis vectors and the residual of the NMF method revealed that a larger number of basis vectors results in a smaller residual. In the experiment of approximating 9 fruit scents, we tried to approximate the ORN responses to target odor by using 10 extracted basis vectors. We found that a larger number of basis vectors results in a higher correlation coefficient.

*Corresponding author: e-mail: nakamoto.t.ab@m.titech.ac.jp

1. Introduction

Some odorants can evoke similar impressions, i.e., the same or a similar sensory impression can be obtained by blending a small number of odor constituents. Thus, an odor representation using all its constituents might be superfluous. From another point of view, their olfactory thresholds are sometimes different from the limits of detection of analytical equipment or artificial sensors such as mass spectrometry. Thus, the distance in artificial sensor space is different from that in biological space. It is necessary to compare sensory space with the artificial sensors' one and to take an appropriate distance measure to fit it so that odor approximation can be established. If we could achieve that, we can establish odor components to represent a wide range of odors, and translate odors into digitized information.

Our group has for years been studying the odor sensing space. We have built a database of several hundred types of essential oil by mass spectrometry (MS).[1–3] We have already reported that basis vectors corresponding to odor components can be extracted from the established database. Our current research also revealed that the accuracy of approximation in sensory evaluation was sometimes low despite the high accuracy in mass spectrum space.

In our latest study, we revealed that the accuracy was improved by the nonnegative matrix factorization (NMF) method,[4,5] based on the distance measure such as the Itakura-Sato (IS) divergence.[5] Because of the reproduction of small mass spectrum peaks despite important contributions, we have already reported that the IS divergence improved the accuracy in approximation of blended essential oils by mixing odor components based on MS data space.[6]

On the other hand, odorant recognitions at olfactory receptors (ORs) are the primary process of biological olfaction. Thus, we focus on the response pattern from a set of olfactory receptor neurons (ORNs) for various odorants instead of the mass spectrum. In this paper, we analyzed the biological sensing space on the basis of response data reported by Hallem and Carlson[7] using the NMF method.

2. Materials and Methods

2.1 *Basis vectors and odor components.*

The outline of the odor approximation method is briefly described in Fig. 1(a). Let odor data be an n-dimensional vector, n be the number of sensors or sensing elements, and m the number of odors. Thus, database V would be an $n \times m$ data matrix. Let WH be approximate matrix, W be an $n \times r$ basis matrix, and H be an $r \times m$ coefficient matrix. V is factorized as

$$V \cong W \times H = WH, \tag{1}$$

where r is the number of basis vectors. Each column of W is the basis vector of the odor component. The extracted basis vectors in W can be used for translating high-dimensional data into low-dimensional space. The obtained coefficients in H represent the mixture composition of the odor components. The odor approximation is performed with approximated basis vectors using existing odors as shown in Fig. 1(b).

Detabase

$$
\begin{bmatrix} v_{1,1} & \cdots & v_{1,m} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,m} \end{bmatrix} \cong
$$

V
$(n \times m)$

Basis vectors

$$
\begin{bmatrix} w_{1,1} & \cdots & w_{1,r} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,r} \end{bmatrix} \times
$$

W
$(n \times r)$

Coefficients

$$
\begin{bmatrix} h_{1,1} & \cdots & h_{1,m} \\ \vdots & \ddots & \vdots \\ h_{r,1} & \cdots & h_{r,m} \end{bmatrix} =
$$

H
$(r \times m)$

Approximated matrix

$$
\begin{bmatrix} wh_{1,1} & \cdots & wh_{1,m} \\ \vdots & \ddots & \vdots \\ wh_{n,1} & \cdots & wh_{n,m} \end{bmatrix}
$$

WH
$(n \times m)$

(a)

Target odor

$$
\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} \cong
$$

t
(n)

(Approximated) Basis vectors

$$
\begin{bmatrix} w'_{1,1} & \cdots & w'_{1,r} \\ \vdots & \ddots & \vdots \\ w'_{n,1} & \cdots & w'_{n,r} \end{bmatrix} \times
$$

W'
$(n \times r)$

Recipe

$$
\begin{bmatrix} s_1 \\ \vdots \\ s_r \end{bmatrix} =
$$

s
(r)

$$
\begin{bmatrix} w's_1 \\ \vdots \\ w's_r \end{bmatrix}
$$

$w's$
(n)

* W' is the approximated W.

(b)

Fig. 1. Extraction of basis vectors and odor approximation by using their basis vectors. (a) Extraction of basis vector W (NMF method etc.). (b) Odor approximation.

2.2 *Biological sensing space*

We have already analyzed the mass spectrum to obtain odor components. However, the obtained odor sensing information should be closer to that of biological olfaction. Thus, we focus on response patterns from 24 ORNs in an insect called *Drosophila*. Ideational representations of the mass spectrum and sensory spaces are shown in Figs. 2(a) and 2(b), respectively. In the mass spectrum space, the distance between odors B and C is short, whereas the corresponding distance in the sensory space is long. It is necessary to compare mass-spectrum data space with biological space.

Hallem and Carlson reported a systematic study of the response data of all ORNs of *Drosophila*.[7] We focused on that insect's olfactory data owing to the following reasons:
· It has already been possible to obtain response data of all types of ORN in several insect species.
· The number of ORN types is much smaller than those of vertebrates. Moreover, their function is simple.
· At the current stage, analysis of insect data is more realistic than that of vertebrate data.

At the current stage, an insect's olfaction is more appropriate for systematic survey than a mammal's one since its database is available. Thus, we applied the NMF method to Hallem and Carlson's database.

The data in ORN[7] consist of excitatory and inhibitory responses of *Drosophila*'s 24 types of ORN to 110 types of odorant chemicals commonly found in nature. The data matrix to be analyzed was constructed by separating the excitatory and inhibitory responses (all elements in this matrix are nonnegative and show magnitude of response strength).
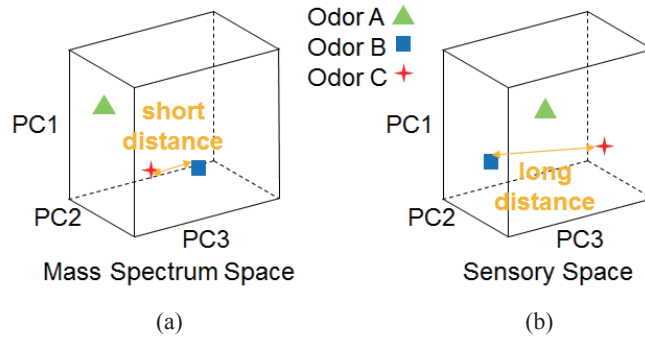
Fig. 2. (Color online) Concepts of mass spectrum and sensory spaces. (a) Mass spectrum space and (b) sensory space.

2.3   *NMF method and odor approximation using odor components*

The outline of NMF can be briefly described as follows. We use the NMF method since the data matrix is factorized under the constraint that any element of W and H must be nonnegative, i.e., all elements must be equal to or greater than zero. Using this dataset, we assumed linear superposition of ORN responses. The basis matrix W and coefficient matrix H can be obtained by iterative calculation as previously reported.[1–3]

In our current research, the nonnegative least-squares method was used for approximation as shown in

$$t \cong W' \times s = w's \qquad (2)$$

where t is the vector of target odor, W' is the approximated W, and s is the mixture composition of the target odor. This method is an optimization method used to solve a linear equation with the constraint that all the elements in the solution vector are nonnegative. In this method, the solution vector s to minimize the distance should be obtained according to

$$I = \min_{s} D_{EU}\,(t|w's) \qquad (\forall s_i \geq 0) \qquad (3a)$$

or

$$I = \min_{s} D_{IS}\,(t|w's) \qquad (\forall s_i \geq 0) \qquad (3b)$$

under the constraint that all the elements in s are nonnegative. $D_{EU}(t|w's)$ and $D_{IS}(t|w's)$ are the distances between t and $w's$ using the Euclidean distance and the IS divergence, respectively. We reported that the minimization method based on IS divergence was also used for the approximation shown in eq. (3b).[6]

3.   Results and Discussion

A small set of odor component is suitable for approximating various odors by using olfactory display because of its hardware cost. Since 10 basis vectors seem to be sufficient for roughly expressing odor space, they were extracted from the matrix by the NMF method given in eq. (1). During the process of exploring basis vectors, the rule for updating matrices in eq. (1) was slightly modified from the original algorithm[4] to achieve good convergence for the insects' response data.

It might be possible that a few tens of basis vectors can approximate odors. To analyze the relationship between the set of extracted basis vectors and the data matrix of olfactory responses, we performed the hierarchical cluster analysis for response data together with basis vectors as shown in Fig. 3. As can be seen, the basis vectors are almost uniformly located according to the spatial distribution of odors. Thus, it might be possible to express a variety of odors using odor components corresponding to the extracted basis vectors.

110 odor chemicals were found to be located spatially as characterized by 24 ORN responses. For example, 1-octen-3-ol's structure (2nd one from the left side) is much different from that of phenethyl alcohol (3rd one from the right side), and the distance between them is long. Conversely, when the chemical structure is similar, the distance is short (for example, nerol and citronellol, 4th and 5th from left of basis vector No. 7). We found that a group of odor compounds with similar structures would form a cluster on this dendrogram, and that the extracted basis vectors were found to be located according to the spatial distribution of odors.

An ORN response pattern of a target odor banana was approximated by using 10 basis vectors obtained with eq. (1). As shown in Fig. 4, the feature of ORN response pattern can be roughly captured using extracted basis vectors. The distance measure used in Fig. 4 was IS divergence.

For the analysis of the relationship between the number of basis vectors and the residual, the basis vectors were extracted from the matrix by the NMF method given in eq. (1). The relationship between the number of basis vectors and the residual after 10,000 iterations is shown in Fig. 5(a). According to Fig. 5(a), it was found that a larger number of basis vectors results in a smaller residual. The distance measure used in Fig. 5 was also IS divergence.

As shown in Fig. 5(b), ORN responses to 9 fruit scents were approximated using various sets of basis vectors. The basis vectors were approximated using the responses of 24 ORNs to 110 odorants and eq. (3b). The ORN responses to fruit scents were also obtained from the literature.[7] The correlation coefficient indicates the correlation between the actual response pattern and the approximated one. It was found that a larger number of basis vectors results in a higher correlation coefficient.
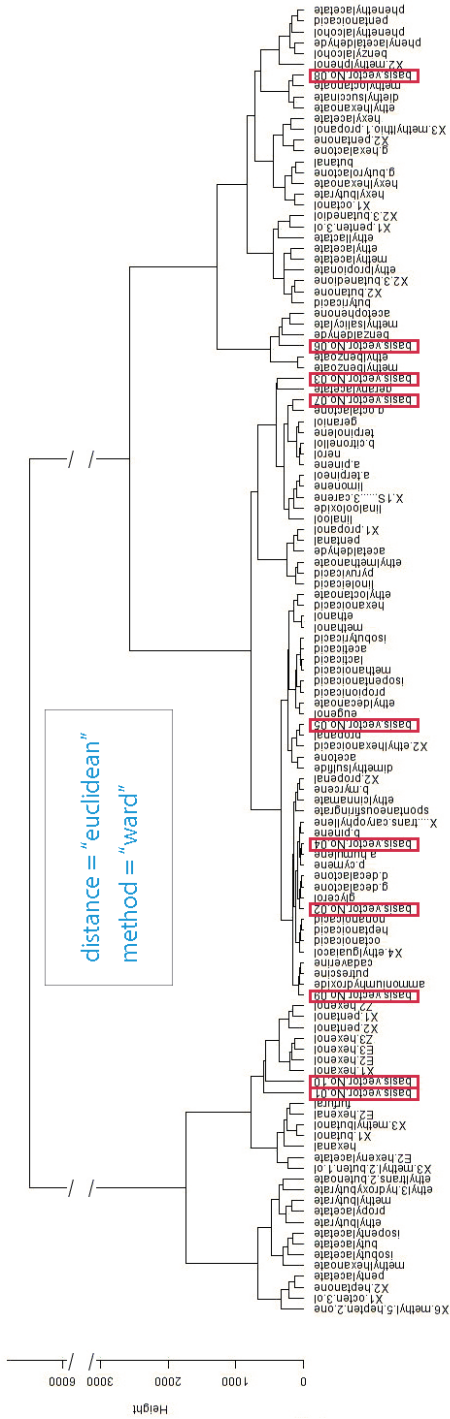
Fig. 3. (Color online) Result of hierarchical cluster analysis of insect's data together with basis vectors of NMF method. Red rectangles indicate extracted basis vectors.
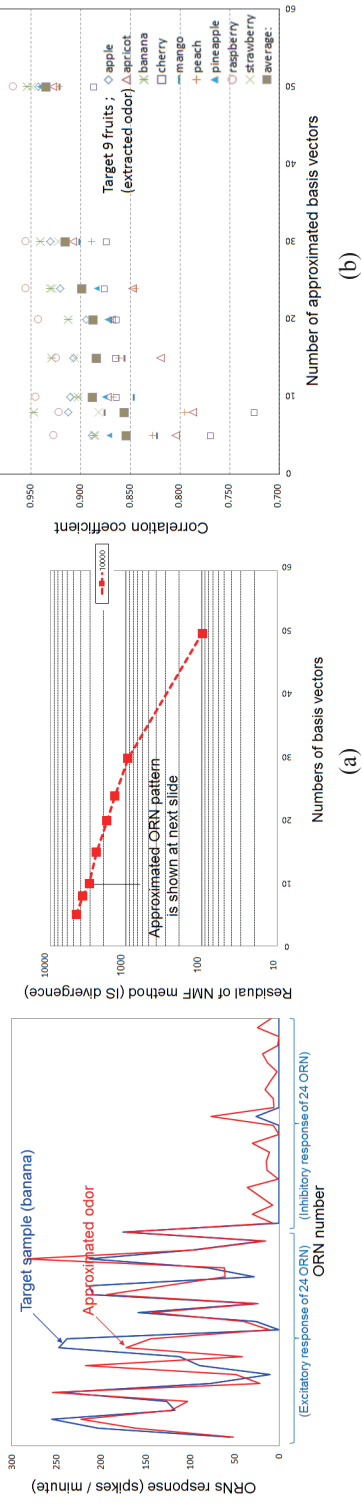

Fig. 4. (left) (Color online) ORNs' response pattern approximated by extracted 10 basis vectors. Distance measure: IS divergence.

Fig. 5. (right) (Color online) (a) Relationship between number of basis vectors and residual after 10,000 iterations. (b) Result of ORNs' response pattern approximations of 9 fruit scents using approximated basis vectors.

4.   Conclusions

An insect's ORN responses were analyzed by the NMF method. Extracted basis vectors were found to be located according to the spatial distribution of odors. The ORN response pattern can be approximated using extracted basis vectors. Although the residual decreases as a function of the number of basis vectors, there is still room for accuracy improvement. Since the solution of NMF is unique, appropriate initial values before iteration calculation are important. The optimal method of initial value setting will enable more stable and better convergence. Moreover, to add to it, the accuracy improvement of ORN response approximation using the basis vectors and sensory evaluation of approximated odors remains to be studied in the future.

References

1   T. Nakamoto and K. Murakami: Virtual Reality Conference (IEEE, 2009) p. 159.
2   T. Nakamoto, M. Ohno and Y. Nihei: IEEE Sens. J. **12** (2012) 3225.
3   Y. Nihei and T. Nakamoto: Sens. J. **13**(2013) 4305.
4   D. D. Lee and H. S. Seung: Nature **401** (1999) 788.
5   H. Sawada: J. IEICE **95** (2012) 829 (in Japanese).
6   Y. Harada, K. Honda and T. Nakamoto: Technical Meeting on Sensors and Micromachines (IEE of Japan, 2013) CHS-13-7, p. 37 (in Japanese).
7   E. A. Hallem and J. R. Carlson: Cell **125** (2006) 143.