Unveiling the Odor Representation in the Inner Brain of *Drosophila* through Compressed Sensing

Kiri Choi[1], **Won Kyu Kim**[1], **and Changbong Hyeon**[1,*]

[1]Korea Institute for Advanced Study, Seoul 02455, Korea; *Correspondence: hyeoncb@kias.re.kr

(Dated: July 10, 2023)

ABSTRACT

The putative dimension of a space spanned by chemical stimuli is deemed enormous; however, when odorant molecules are bound to a finite number of receptor types and their information is transmitted and projected to a perceptual odor space in the brain, a substantial reduction in dimensionality is made. Compressed sensing (CS) is an algorithm that enables the recovery of high dimensional signals from low dimensional data compressed through a sensing matrix when the representation of such signals is sufficiently sparse in the vector space. Here, we use the recent *Drosophila* connectomics data to show that the sensing matrix defined from two interneuronal synaptic interfaces effectively fulfills the prerequisites for CS to work. We demonstrate the activity profile of projection neurons (PNs) can be faithfully recovered from a lower dimensional response profile of mushroom body output neurons (MBONs) constructed using the experimentally measured electrophysiological responses to a wide range of odorants. By leveraging the residuals calculated between the measured and the predicted MBON responses, we visualize the perceptual odor space in terms of residual spectrum and discuss the differentiability of a given odor. The simultaneous introduction of many different odorants saturates the activity profile of PNs, significantly degrading the capacity of signal recovery, resulting in a perceptual state analogous to "olfactory white." Our study that applies CS to the connectomics data of the *Drosophila* olfactory system provides novel and quantitative insights into the structure of perceptual odor space and odor representation in the brain.

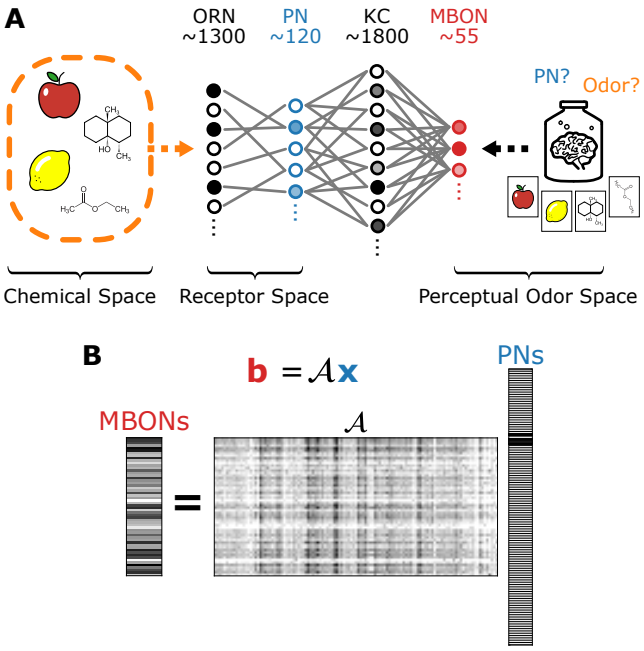


Figure 1: **Olfactory processes in *Drosophila*.** (A) An odor represented in the chemical space is encoded by a sparse set of ORs/PNs in the receptor space. Highly integrated olfactory signals in the perceptual odor space are deciphered at higher cortical areas, which infers the activity profiles of uPNs (blue) and hence the olfactory code from the relatively low dimensional and randomized information integrated into MBONs (red). (B) Our problem is formulated to solve for n-dimensional sparse olfactory codes ($\mathbf{x}$) encoded by PNs from a set of underdetermined linear equations in the form of $\mathbf{b} = \mathcal{A}\mathbf{x}$, where $\mathbf{b}$ and $\mathcal{A}$ represent the $m(< n)$-dimensional response profile of MBONs and the $m \times n$ sensing matrix, respectively.

INTRODUCTION

Olfaction is a sensory process of detecting chemical stimuli in the environment. Despite its fundamental importance, still left ambiguous are basic quantities associated with the understanding of olfaction, such as the number of differentiable odors and the dimension of odor percepts [1–3]. This is in stark contrast to other sensory processes such as vision, which is quantitatively understood as the combined responses of distinct types of photoreceptors to visual signals continuous in the spectral domain. The odor identification and perception are still deemed highly subjective and context-dependent, varying from one individual to another [4]. However, the conditions for understanding the insect olfactory system have greatly been improved in recent years, thanks to the three-dimensional map of neuronal morphology and the synaptic connectivity reconstructed from high-resolution electron microscopy (EM) images, such as *Drosophila* hemibrain dataset [5], the olfaction-specific FAFB dataset [6], and more recently the full connectome of the *Drosophila* larva [7]. By making the neural circuitry of the insect brain accessible at synapse resolution, these connectomics datasets open a new avenue to a quantitative understanding of the principles behind olfactory processing.

As illustrated in Figure 1A, olfactory signaling starts when an odorant binds to olfactory receptors (ORs) in the olfactory receptor neurons (ORNs). Each ORN expresses a specific type of OR, exhibiting homotype-dependent responses to input odorants. The same type of ORNs converges to a particular set of uniglomerular PNs (uPNs) forming a glomerulus in the antennal lobe (AL) [8]. The signals converged and sorted at uPNs transfer to the higher olfactory centers such as the mushroom body (MB) calyx and the lateral horn (LH), where synaptic wirings with higher order neurons integrate the signal. Specifically, extensive and seemingly randomized signal integration occurs at the synaptic interface between PNs and the Kenyon cells (KCs) [9, 10], which passes over to MBONs whose dimension is smaller than that of PNs. In LH, the compartmentalized spatial organization of PNs translates to a more stereotyped signal integration by lateral horn neurons (LHNs) [11–14].

For given odor representation at MBONs, there may exist a neural process analogous to solving a constrained optimization problem that enables higher cortical areas to know the identity of the odor in the form of "odor code", defined as a combinatorial expression of uPN activities following the concept of combinatorial coding [10, 15–19] (Figure 1A). Among several candidate algorithms, compressed sensing (CS), which determines the sparsest solution of a set of underdetermined linear equations, has previously been applied to the *Drosophila* olfactory system as a means to identify the chemical composition of an odor from activity profiles (i.e., neuronal firing rates) of ORNs, PNs, or KCs [17, 20–23]. Provided that the chemical composition of an odor represented in the chemical feature space is sparse [24, 25], CS can reliably uncover the odor identity. CS has been applied to the olfactory sensing in the scope of the odorant-OR interface [17, 22, 23] and between odorants and KCs [19, 21]. The previous studies using CS on olfaction have offered valuable insights into the odor codes and odor identification process; yet, they are largely limited to the relationship between the chemical feature space and the receptor space.

By adopting the neural circuit of the *Drosophila* olfactory system from the connectomics data, we investigate (i) whether the idea of CS [10] enables an accurate reconstruction of odor code from the highly integrated and compressed olfactory information at MBONs (Figure 1A, B) and (ii) whether the perception of a given odorant(s) is unambiguously discernible from other odors in the perceptual odor space. We test the capacity of CS in recovering uPN activities from MBON response profiles reconstructed from the electrophysiological dose-response experiments [25, 26]. In particular, we calculate the spectrum of residuals to visualize the perceptual odor space by comparing the reconstructed odor response profiles to those projected onto other olfactory stimuli (see Figure 2A). We explore how the perceptual odor space changes when multiple odorants are mixed, demonstrating that the simultaneous introduction of diverse odorants significantly deteriorates the capacity of CS, an observation in line with the notion of "olfactory white." We also show that a significant variation in concentration can yield a large deviation in the odor code and alter the extent of odor identifiability. By employing the CS algorithm as a means to probe the olfactory processing in the inner brain of *Drosophila*, our study presents a novel perspective on odor perception while demonstrating the functional benefits of randomized wiring and sparse coding.

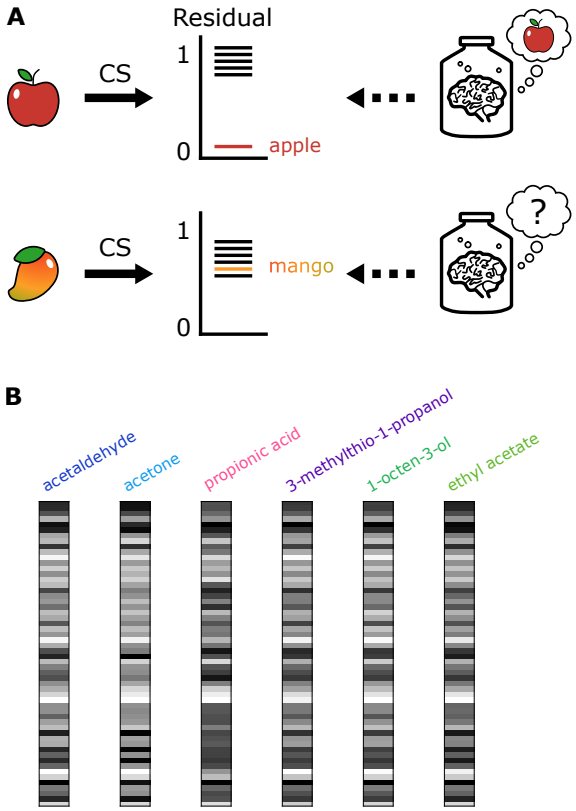


Figure 2: **Odor representation and perceptual odor space.** (A) A putative structure of the perceptual odor space. The gap between the self-residual and cross-residuals decides the differentiability of a given odor (e.g., apple, mango in the illustration) from others. (B) The MBON response profiles of select odorants (normalized for better visualization). Despite functional and qualitative distinctions between odorants, response profiles can look similar. For the complete MBON response profiles of all the tested odorants, see Supplementary Figure S3.

RESULTS

The capacity of compressed sensing in signal recovery

We first demonstrate that CS can reliably recover the uPN activity profile from the MBON response profile through the sensing matrix defined by the product of two synaptic connectivity matrices (see Supplementary Figures S1 and S2, and Methods for a detailed discussion on CS). Consider a situation depicted in Figure 1A, where *Drosophila* whiffs the distinctive odor of an apple. The apple odor in the chemical space will be encoded into ORN/uPN activity profiles. Before an olfactory signal reaches the higher cortical area, the signal is integrated as it passes through higher-order olfactory neurons such as KCs and MBONs. Then, is it possible for the higher cortical area to infer a high-dimensional uPN activity from a low-

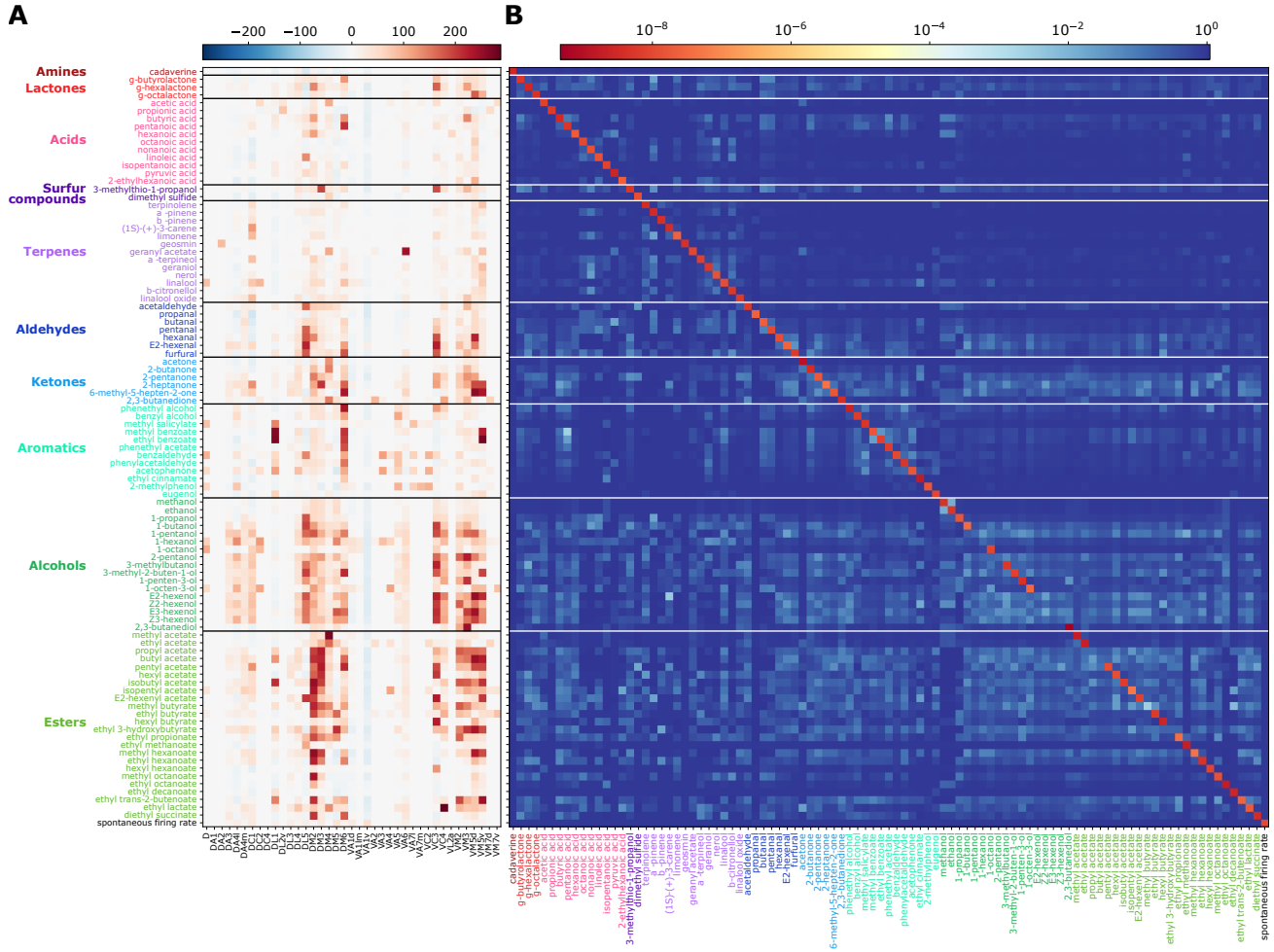


Figure 3: **uPN activity profiles and the residuals.** (A) The electrophysiological recordings of 39 homotypes to 96 odorants (+ basal response labeled as 'spontaneous firing rate') where at least a single homotype exhibited a strong response (≥ 40 spikes/s). The labels are color-coded based on the same odor categorization used by Hallem *et al.* [25] (Dark red: amines, red: lactones, pink: acids, purple: sulfur compounds, violet: terpenes, dark blue: aldehydes, blue: ketones, emerald: aromatics, green: alcohols, olive: esters) (B) The residual $r_{\alpha|\beta}$ calculated for an input odorant α against other odorants β denoted in the horizontal axis.

dimensional MBON response profile and identify the odor code as an apple? It is not straightforward to gauge the perceptual differences between odors by simply comparing two MBON response profiles represented in 56-dimensional vectors (see Figure 2B, Supplementary Figure S3, and Supplementary Information). Therefore, given the MBON response profile, we leverage CS to solve the inverse problem of finding the sparest representation of the odor in uPN activities instead. Ideally, the MBON activity expected from the solution of CS processed through the sensing matrix should reproduce what was given. Provided that the odor code of an apple is well-discernible from that of other odors, the odor representation of an apple at MBONs should be unique and discernible from others as well.

In the technical terms of CS, an MBON response profile and a uPN activity profile are represented as the vectors $\mathbf{b}$ and $\mathbf{x}$, respectively (Figure 1B and Supplementary Figure S3). We build uPN activities ($\mathbf{x}^{meas,\alpha}$) from the electrophysiological recordings [25, 26] for various odorants or odorant mixtures α and calculate the corresponding MBON response profile via the relation $\mathbf{b}^{meas,\alpha} = \mathcal{A}\mathbf{x}^{meas,\alpha}$. CS uses $\mathbf{b}^{meas,\alpha}$ and $\mathcal{A}$ to determine the sparest uPN representation for α ($\hat{\mathbf{x}}^{\alpha}$), from which we calculate the "residual" denoted by $r_{\alpha|\beta}$ (see Methods and Extended Figure 2), defined between the measurement ($\mathbf{b}^{meas,\alpha}$) and the calculation ($\mathcal{A}\hat{\mathbf{x}}^{\alpha}$) projected on itself ($\mathcal{A}\delta_{\alpha}(\hat{\mathbf{x}^{\alpha}})$) or other odor ($\mathcal{A}\delta_{\beta}(\hat{\mathbf{x}^{\alpha}})$). The spectrum of residuals ($\{r_{\alpha|\beta}\}$), especially the gap between the self-residual ($r_{\alpha|\alpha}$) and the cross-residuals ($r_{\alpha|\beta}$, $\beta \neq \alpha$), provides an insight into the differentiability of the odor α in the perceptual odor space (Figure 2A). If the self-residual is small and well separated from all other cross-residuals ($r_{\alpha|\alpha} \ll r_{\alpha|\beta}$ for all $\beta \neq \alpha$), then the representation of odor α in the perceptual space is unique and easily discernible from others (Figure 2A, top). In contrast, if the self-residual is comparable to other cross-residuals ($r_{\alpha|\alpha} \simeq r_{\alpha|\beta}$), it is

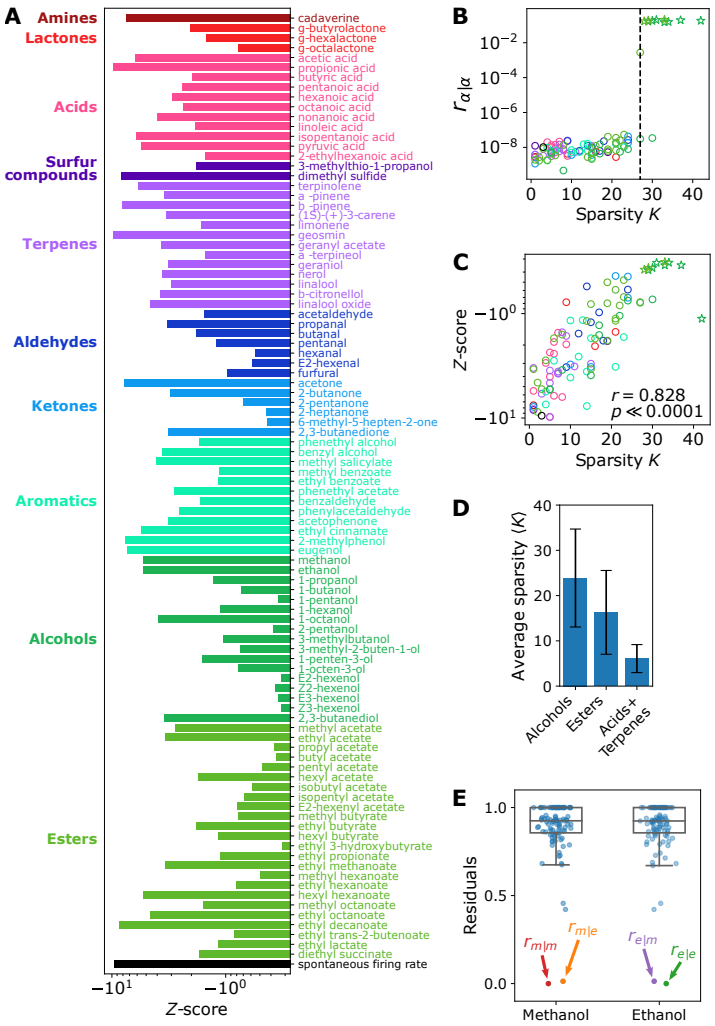


Figure 4: **Z-score and sparsity K.** (A) Z-scores of the self-residuals for 96 odorants calculated from Figure 3B. (B) Self-residual $r_{\alpha|\alpha}$ versus the sparsity of $\mathbf{x}^{meas,\alpha}$. The odorants whose self-residual ($r_{\alpha|\alpha}$) is not minimal in the spectrum of residuals are marked with the star symbol. The black dashed line denotes the midpoint sparsity ($K = 27$) obtained by fitting the data to a logistic function. (C) Z-scores of $r_{\alpha|\alpha}$ versus the sparsity (K) of $\mathbf{x}^{meas,\alpha}$. (D) The average sparsity of the odor response profile for odorants that displayed strong responses ($\geq$ 40 spikes/s) under the specified functional group. (E) Distributions of the residuals for $\alpha = \mathrm{MeOH}$ and $\alpha = \mathrm{EtOH}$ ($r_{m|m} = r_{\mathrm{MeOH|MeOH}}$, $r_{m|e} = r_{\mathrm{MeOH|EtOH}}$, $r_{e|e} = r_{\mathrm{EtOH|EtOH}}$, and $r_{e|m} = r_{\mathrm{EtOH|MeOH}}$).

difficult to differentiate the odor α from other odors, implying that the perceptual odor space of α is not well isolated (Figure 2A, bottom). The spectrum of residuals $\{r_{\alpha|\beta}\}$, which offers glimpses into the high dimensional perceptual odor space, allows us to assess the specificity of the odor α among other odors.

Embracing the criterion by Seki *et al.*, who used either 30 spikes/s or 50 spikes/s as the detection threshold to demonstrate that a strong response to an odor is sparse [26], we define the strong response as the one evoking more than 40 spikes/s, analogous to the "winner takes all" (WTA) approach [19], and take into account only the odorants that elicit a strong response to at least one glomerular homotype in the experiments. The combined dataset contains the recordings from 39 glomeruli, which correspond to ~ 80 uPNs (Figure 3A), in response to 96 different odorants encompassing diverse odor types that feature different functional groups and behavioral responses (Extended Table 1). Here, we assume that the activation of neurons is linearly correlated with equal gain in response to the odorants [27] and that uPNs extending from the same glomerulus receive identical input. Furthermore, we also assume that ~ 40 uPNs out of 119 uPNs included in our circuit, which are not part of the measured 39 glomeruli, do not respond to the tested odorants.

Our CS framework correctly identifies 86 odorants out of 96 odorants ($\sim 90\%$) (see Figure 3B and Extended Figure 1), indicating that $\mathbf{b}^{meas,\alpha} \approx \mathcal{A}\hat{\mathbf{x}}^{\alpha}$ (and $\mathbf{x}^{meas,\alpha} \approx \hat{\mathbf{x}}^{\alpha}$), with $r_{\alpha|\alpha}$ being minimal among all the residuals ($\{r_{\alpha|\beta}\}$). The 10 odorants that failed our test include 1-hexanol, 1-pentanol, 2-pentanol, cis-2-Hexen-1-ol (Z2-hexenol), trans-2-hexen-1-ol (E2-hexenol), cis-3-Hexen-1-ol (Z3-hexenol), trans-3-hexen-1-ol (E3-hexenol), propyl acetate, ethyl 3-hydroxybutyrate, and butyl acetate, all of which are either alcohols or esters. For the statistical analysis of the residual spectrum for an odorant α ($\{r_{\alpha|\beta}\}$), we calculate the Z-score of the self-residual ($r_{\alpha|\alpha}$) (see Methods). Many odorants exhibit significantly negative Z-scores, indicative of a large gap between $r_{\alpha|\alpha}$ from $r_{\alpha|\beta}$ and a well-partitioned perceptual odor space for α (Figure 4A).

We find that the successful identification of odorant α hinges on the sparsity (K) of its input uPN activity profiles $\mathbf{x}^{meas,\alpha}$ (Figure 4B). An activity profile with a large K, which is found to correlate with the Z-score (Figure 4C), gives rise to a large self-residual ($r_{\alpha|\alpha}$) resulting in a Z-score with $Z \lesssim 0$, which implies a non-optimal solution of CS. For accurate odor identification, the sparsity should be smaller than $K = 27$, a number close to the theoretical limit where our CS setup with the predetermined dimensions might have the most difficulty in solving for the sparest solution (see Supplementary Information and Figure S4A).

Odorants with certain functional groups, such as acids, terpenes, and aromatics, are characterized by Z-scores with $Z \ll 0$ (see Figures 3B and 4), implying that the uPN activation profiles for these odorants are unique at least among the odorants we tested, and hence the corresponding perceptual odor space can be unambiguously discerned from others. In contrast, alcohols and esters elicit stronger responses over many uPNs [28], resulting in $Z \lesssim 0$ (see Figures 3 and 4A). The average sparsity ($\langle K \rangle$) of the activity profiles for alcohols and esters is greater than that of a combination of acids and terpenes by $3 - 4$ fold (Figure 4D). The considerable difference in $\langle K \rangle$ likely leads to the poor identifiability of many alcohols and esters. This finding is consistent with a previous study suggesting that ORs encoding alcohols and esters are broadly tuned, while another set of ORs that encode acids and amines are narrowly tuned [29]. The relatively poor odor identifiability of alcohols and esters, revealed from our CS framework, might also be

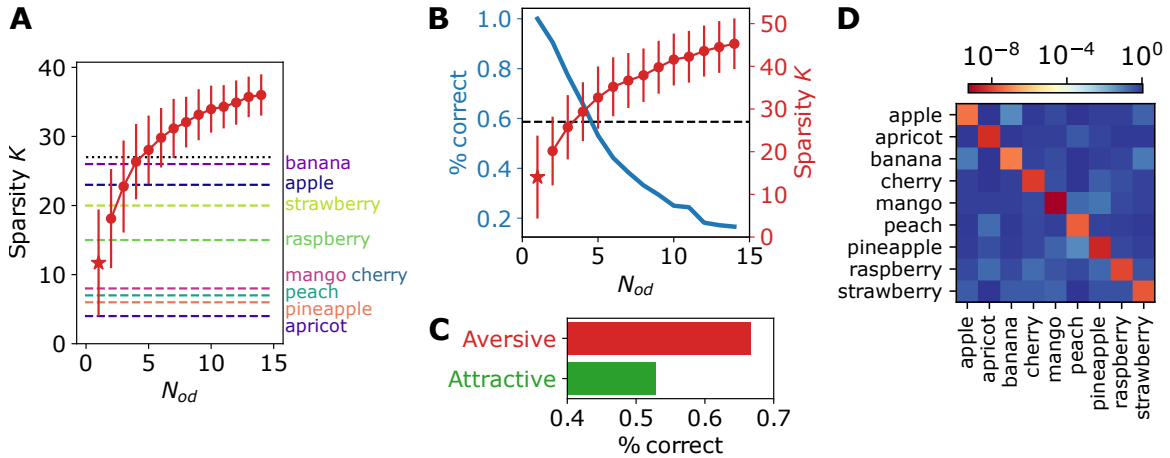


Figure 5: **Analysis of odor mixtures.** (A) The average sparsity of randomly sampled odor mixture with different sample size N_{od} (red dots) and the sparsity of $\mathbf{x}^{meas,\alpha}$ when $\alpha =$ natural mixtures (dashed line). The sparsities are counted based on the 23 glomeruli whose responses were recorded in both the single odorant experiment and the fruit odor experiment for consistency. The red star denotes the average sparsity for single ($N_{od} = 1$) odorant exposure. The error bars indicate the standard deviation. The dotted black line corresponds to $K = 27$ in Figure 4B. (B) The input signal recovery (blue line) and the corresponding average sparsity (red dots) of a random mixture of odorants with different sample size N_{od}. The output is from the combined dataset measuring 39 glomeruli against 86 odorants whose self-residual ($r_{\alpha|\alpha}$) was minimal. The red star denotes the average sparsity for single ($N_{od} = 1$) odorant exposure. The red errorbar indicates the standard deviation. The dashed line corresponds to the midpoint sparsity ($K = 27$). (C) The probability of successful input signal recovery for the mixture consisting of two attractive odors (green) or two aversive odors (red). (D) The matrix of the residuals, $r_{\alpha|\beta}$, calculated based on fruit odors. Each row depicts the spectrum of the residuals of an odor input α with respect to other fruit odors β ($\beta =$ apple, apricot, ...) denoted in the horizontal axis.

related to the fact that the information necessary to identify alcohols and esters is processed through channels other than MBONs. Specifically, stereotyped spatial segregation of fruit-odor encoding PNs has been observed [11], which leads to stereotyped connectivity to LHNs [14, 30], suggestive of the existence of labeled lines [6, 12, 14, 26, 31, 32]. The odor identification and the subsequent decision-making process for alcohols and esters may heavily rely on the co-processing of information transmitted through both KCs and LHNs, as a sizable portion of LH projections converge with the downstream projection from MB calyx [6, 33, 34].

Still, esters such as ethyl methanoate and ethyl decanoate are well discernible in the perceptual space. The same holds for methanol and ethanol, except when compared to each other. The odor identifiability assessed from our CS framework suggests that *Drosophila* may not be able to tell the difference between methanol and ethanol [25], as the corresponding values of self- and cross-residuals are almost identical to each other ($r_{m|m} \approx r_{m|e}$ and $r_{e|e} \approx r_{e|m}$. See Figure 4E).

Perceptual space of odor mixture and olfactory white

In natural environments, most odors originate from complex mixtures of odorants. In humans, it is known that exposure to multiple odorants leads to the loss of the odor characterization of individual components [35–37], with a mixture consisting of an even greater number of odorants leading to a perceptual state called "olfactory white" [38]. This loss of odor differentiability is analogous to the loss in CS capacity, leading to a question of how simultaneous exposure to multiple odorants affects our CS framework. To test this hypothesis, we examine our framework against (i) random mixtures of a predetermined number of odorants and (ii) natural odor mixtures of fruit extract.

When multiple odorants are randomly mixed, how does the brain perceive the odor? At the glomerular level, it has been suggested that an odorant mixture elicits glomerular activities [39] that may or may not be identical to the sum of the activities of individual glomeruli [40–42]. Competitive binding of odorants to receptors is expected [42], and some odorants may be more dominant than others depending on the chemical and perceptual composition of the mixture [27]. Here we try to assess the odor identifiability from the vector $\mathbf{x}^{meas,\alpha}$ encoding a natural odor whose precise chemical composition is unknown. First, we estimate the sparsity of the uPN activity profile associated with natural odors since the challenges encountered by the olfactory system translate to the difficulty of signal recovery via CS. We compare the results from single odorants and fruit extracts, the latter of which are available from Hallem *et al.* in the form of electrophysiological recordings on 23 glomeruli [25]. To make a fair comparison of fruit extracts against the mixture of single odorants, we analyze the responses to the single odorants using only 23 glomeruli, which are extended to the 47 uPNs, as in the actual measurement (see Figure 5A). The chemical composition of fruit odors is expected to be diverse, and we observe a large variation in the sparsity among fruits. The sparsities of some fruits (e.g., banana) are high, to the point that their odor identifiability through CS is anticipated to be low. There is a chance that the spar-

sity reported here is underestimated since other sets of glomeruli not recorded in the experiment may have activated. Furthermore, a natural odor, whose chemical composition is complex in reality, could display a small K when it is encoded by a small set of broadly-tuned ORs [43].

For random odorant mixtures, we directly compared vectors $\hat{\mathbf{x}}$ and $\mathbf{x}^{meas}$ to assess the accuracy of the signal recovery instead of calculating residuals (see Figure 5B and Methods) so that we define the recovery to be successful if the cosine distance between the two vectors is small. From the 86 odorants that the CS framework accurately identified, we randomly sampled the odorants to create 500 random mixtures with size N_{od}, linearly combined the uPN activity profiles of individual odorants in each mixture, and performed CS to infer $\hat{\mathbf{x}}$. As expected, we observe a steady decline in the chance of successful recovery of the original signal $\mathbf{x}^{meas}$ as the number of odorants composing a mixture (N_{od}) increases, with the chance dropping below 20 % for $N_{od} > 10$ (or $K > 40$) (see Figure 5B). When many odorants are mixed at comparable concentrations, the signal recovery becomes less effective under our CS framework.

However, the nature of the random odorant mixture can vary when the mixture is restricted to a specific valence. We select nine attractive odorants and seven aversive odorants, all of which are well characterized in terms of *Drosophila* behavioral response [6, 14, 31, 32, 44–47] (Extended Table 1) and create combinations of attractive or aversive odors by sampling two odorants from each category. We find that the capacity for signal recovery is significantly poor for the mixture of attractive odors compared to the mixture of aversive odors (Figure 5C). This is due to the generally poor identifiability of alcohols and esters, which many attractive odorants are a part of (see Figure 4A). While our result may not be directly translatable to the odor perception of *Drosophila in vivo*, the functional consequence of loss in the identifiability of an odor mixture seems analogous to the notion of "olfactory white."

In the case of natural mixtures (fruit extracts), we associated an activity profile for a fruit odor with a percept, such that α designates a type of fruit (e.g., apple) instead of an odorant. The measurements were made in response to the concentrated extracts from nine fruits (apple, apricot, banana, cherry, mango, peach, pineapple, raspberry, and strawberry) diluted to 10^{-2} volume per volume (v/v) with H_2O. Based on the results from random odorant mixtures, the recovery of fruits such as apples and bananas whose uPN activity profile has high sparsity does not seem guaranteed (Figure 5A). However, the spectrum of residuals is still specific enough to resolve the corresponding odors among a list of fruits, suggesting that their perceptual odor spaces in the brain of *Drosophila* are still well separated from other fruits (see Figure 5D, Supplementary Figures S5 and S6).

Concentration-dependent odor perception

So far, our analysis has been conducted using electrophysiological recordings of odorants at a relatively high concentration. Both the recordings provided by Hallem and Carlson [25] and Seki *et al.* [26] employed 10^{-2} dilution with either H_2O, paraffin oil, or mineral oil (except for geosmin, which used 10^{-3} dilution) [26].

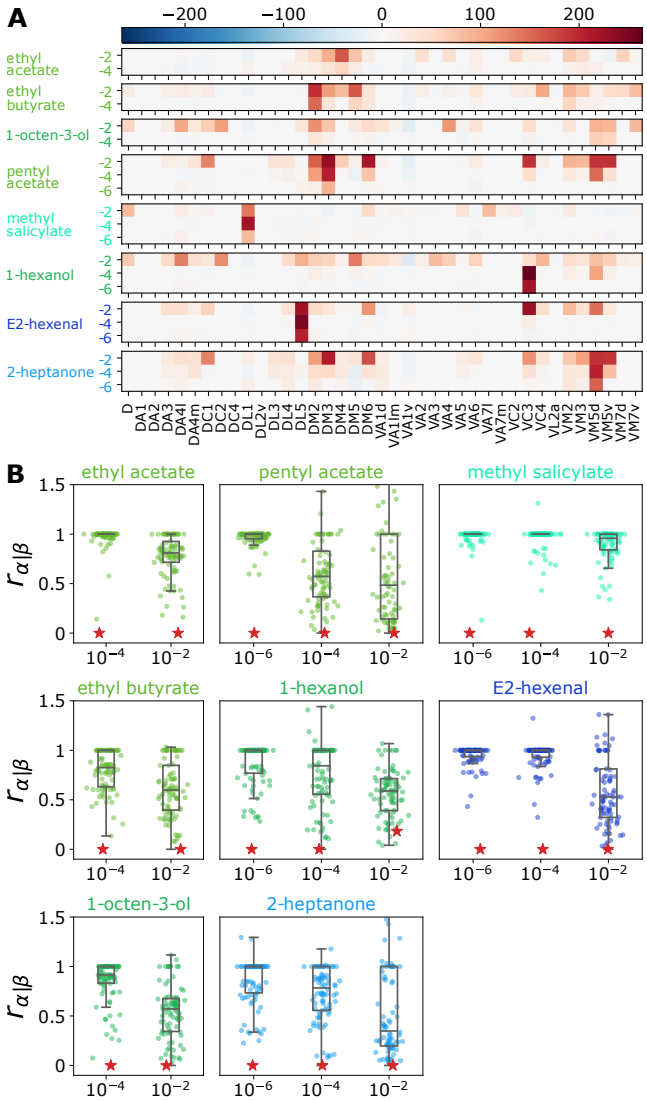


Figure 6: **Concentration-dependent perception of odor.** (A) Electrophysiological recordings of uPN activity profiles in response to select odorants measured at different dilution levels (-2, -4, and -6 on the vertical axes signify the dilutions of 10^{-2}, 10^{-4}, and 10^{-6}, respectively). Only those that elicited a strong response to at least one homotype are plotted. (B) The distributions of $r_{\alpha|\beta}$ for odorants at different dilution levels. Red stars denote the self-residual $r_{\alpha|\alpha}$.

It is, however, known that the activity profile of ORNs and PNs are concentration-dependent [18, 25], and so is the observed hedonic valence of *Drosophila* to a given odorant [45, 48]. Similar concentration-dependent behavioral responses have been observed in *C. elegans* [49].

Figure 6A summarizes the electrophysiological responses of glomeruli to a small subset of odorants at varying dilution levels (10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}) [25]. Some odorants evoke strong responses at low dilutions while others do not, consistent with previous studies on odorant-dependent odor detection thresholds [50]. Odorants at unusually low concentrations may still be perceivable

through a combination of specialized olfactory processing [51, 52]; however, for consistency, we limit our study to dilution levels where there is an activation from at least a single neuron (see Figure 6A and Supplementary Figure S7). All eight odorants failed to evoke strong responses at 10^{-8} dilutions.

When performing CS, we included all previously tested odorants, which were measured at 10^{-2} dilution while replacing the odorant α with different dilution levels. 1-hexanol, an odorant that our CS framework could not correctly identify at a high concentration, became identifiable at lower concentrations, presenting a minimal self-residual (see Figure 6B). A significant reduction in the sparsity K at a lower odorant concentration (Supplementary Figure S7) helps CS accurately recover the uPN activity profile. Our finding also translates to natural mixtures. The high identifiability of fruit odor at 10^{-2} dilution drops significantly for undiluted fruit extracts (see Supplementary Figure S8).

Significant changes are observed for the residual spectra of the same odorant at different concentrations (see Figure 6B). For example, the residual distributions of pentyl acetate and trans-2-hexenal exhibit a drastic change from 10^{-6} to 10^{-4} and from 10^{-4} and 10^{-2} dilutions, respectively. Physiologically, a drastic shift in the residual distributions implies considerable changes in uPN activity and the corresponding MBON response profiles (see Supplementary Figure S9). We believe this signifies an alteration of the odor percept in the perceptual space to the point that the odorant under two concentrations may be perceived as different odors. For ethyl butyrate, our calculation shows a concentration-dependent shift in the residual distribution, explaining the conflicting information on the *Drosophila* behavioral responses reported in the literature [6, 45, 53]. These results are reminiscent of the changes in the qualitative properties reported in human odor characterization experiments.

In conclusion, the sparse encoding of odor to the receptor space is a prerequisite for odor identifiability under our CS framework, and an odor under two vastly different concentrations may occupy disparate perceptual spaces, leading to a different odor percept.

DISCUSSIONS

The neurons and the synaptic connectivity associated with the *Drosophila* olfactory system are akin to the nodes and weight matrices of artificial neural networks (ANN). Whereas an ANN dynamically trains the weight matrix to gain functionality, our system operates on a pre-configured synaptic weight matrix. The synaptic characteristics observed from the *Drosophila* olfactory system, such as the seemingly randomized synaptic interface between uPNs and KCs (Supplementary Figure S1) and the sparsity of uPN activity profiles (Figure 3A), are reminiscent of the prerequisite for CS. Hence, we have applied CS to the physiological circuits of the *Drosophila* olfactory system to explore whether or not it is capable of inferring the identity of the sensed odorant from its representation in the inner brain. By leveraging the relations of neural codes in the receptor space to the perceptual odor space represented by the MBON response profiles, we study the differentiability of an odor stimulus in the perceptual odor space and provide a conceptual explanation of perceptual phenomena such as olfactory white and concentration-dependent odor perception.

As previous studies have acknowledged, whether or not CS is implemented in the *Drosophila* brain is unknown [17, 54]. There is no direct evidence that the higher cortical area effectively leverages the randomization of the PN-KC interface and that there exists a necessary circuitry to perform l_1 norm minimization, although there are suggestions for the implementation of sparse coding in real organisms [55, 56]. Nonetheless, it is fascinating to discover that the *Drosophila* olfactory system exhibits the structural prerequisites for CS. We hypothesize that for the downstream neural circuitry beyond the odorant-OR interface, the compression of the information for chemical stimuli through the effectively randomized sensing matrix is beneficial for conserving signal integrity while reducing the bandwidth and resources necessary for signal transmission to the higher cortical area. It is also interesting to note that a randomly pre-wired network has shown to be more cost-effective with improved accuracy and network trainability in the recent computational development in ANN [57].

While not explicitly considered in this study, other neuronal architectures of the olfactory system are known to play critical roles in signal processing. The most notable are the peripheral interactions [42] present in the second and third-order olfactory neurons. By suppressing the sensitivity of PNs to ORN inputs, lateral inhibitions [58] can allow the encoding of a broader range of concentrations [43], sparsifying the olfactory signal. Mixing highly overlapping odorants that activate similar PNs is known to promote lateral inhibition [59], which could enhance the chance of odor identification even if the uPN activity profile is over-saturated in our framework. Additionally, noise filtration is implemented through inhibitory circuits composed of various antennal lobe local neurons (LNs) [60–62], GABAergic inhibitory PNs (iPNs) [13, 30], anterior paired lateral neurons (APLs) [6, 63, 64], etc. The regulations of olfactory signals across glomeruli by LNs or KCs by APLs lead to noise reduction and sparsification of odor representation [24, 58], both of which should improve the reliability of our CS framework. While we find that the 86 odorants maintain a respectable level of recovery even in the presence of a significant amount of noise (see Supplementary Figure S10), the reduced noise can further sparsify the vector $\mathbf{x}^{meas}$ and ease recovery and identification of the correct activity profile in theory, especially for an odor mixture where a low sparsity of $\mathbf{x}^{meas}$ is not guaranteed. Moreover, ORN-PN sensitivities are glomerulus-specific [65], indicating differential gain control. The ORN response to an odor is nonlinear [18, 25], and a more sophisticated and physiologically correct model for the firing rate could have been used [17]. Nonlinearity prevalent in the olfactory system [18, 43, 60] plays a significant role in olfactory processing [17, 22, 63], and many of these regulatory circuits display odor-type specificity [13, 43].

The temporal dynamics of the signal are critical for olfactory processing as well. For odor perception, time-varying changes in odor concentration play a pivotal role in odor identification, decision-making, and odor-guided navigation [16, 66, 67]. Other interesting

psychophysical phenomena such as olfactory fatigue (peripheral olfactory adaptation) [68], require temporal information to process. Additionally, the temporal dynamics are closely related to the primacy coding hypothesis, which predicts earlier olfactory signals to govern the overall odor identity [16], thereby significantly reducing the difficulty of signal recovery at higher concentrations under our CS framework if implemented. Unfortunately, we could not implement primacy coding due to the limitation in data, although it may be possible to infer the temporal characteristics from the concentration-dependent changes in the uPN activity profile as they highlight ORs most sensitive to a given odorant which should correlate with the order of activity. While we believe primacy coding is an effective principle behind concentration-invariant odor identification, we suspect that the less sensitive and therefore delayed ORN responses may also contribute to the odor perception as characterized by the concentration-dependent quality changes observed in humans.

Lastly, our framework does not consider the anatomy and physiology behind synapses and neurons, which have been shown to contribute to complex olfactory computation such as independent gain control [69]. The locations of presynaptic terminals can be crucial in multimodal sensory processing, and the differentiation between axo-dendritic and axo-axonic connections is necessary for spatiotemporal dynamics [70].

In this work, we explored the concentration-dependent changes to projected perceptual space by entirely attributing it to the changes in the odor codes ($\mathbf{x}^{meas}$). Regarding the concentration-dependent valence, the aversion at high concentration is mediated by the signal through LH that defines the innate responses, and the innate responses are in turn modulated by the output from MB calyx [26]. In light of classical conditioning experiments on *Drosophila* [71, 72], which demonstrated the dramatic change in the *Drosophila* behavior in response to electric shock, it can be surmised that behavioral responses generally depend on both the sensory processes of odor identification and the learned responses. We assume that modification in behavioral response (or decision-making) may result from the multimodal integration of various sensory inputs not covered in this study.

Taken together, we demonstrate the capacity of the CS framework to recover high-dimensional uPN activity profiles from low-dimensional *barcodes* at MBONs in the *Drosophila* olfactory system. In fact, partial information on the MBON response profile still suffices to recover the corresponding uPN activity profile for many odorants (see Supplementary Information), and there certainly is much room for more efficient data compression and retrieval. It will still be of great interest to explore the effects of incorporating the various neuronal features listed above into the basic structure of our CS framework.

METHODS

Data preparation

In this study, we used the hemibrain dataset [5] (hemibrain v1.2.1). When querying the neurons, we chose only PNs, KCs, and MBONs that make at least one synaptic connection. We then collected a total of 119 uPNs synapsing with 1776 KCs, which in turn synapse with 56 MBONs. As for the glomerular labels, we used the one supplied by the hemibrain dataset, except when a uPN was associated with multiple glomeruli. In such a case, we referred to the labeling from the FAFB dataset [6], which provided their version of the glomerular label, along with the corresponding neuron ID in the hemibrain dataset. Additionally, we updated several labels following the recent community agreement to rename three glomeruli with conflicting glomerulus labels. As a result, homotypes VC3l, VC3m, and VC5 are changed to VC3, VC5, and VM6, respectively [34], which have been manually incorporated into the dataset. In this study, we designate the PN type (i.e., homotype) of a PN by the glomerulus innervated by the dendrites. Therefore, the homotype is uniquely defined for uPNs. Supplementary Figures S1 and S2 show the connectivity matrices used in this study.

The odor response profile was generated by merging two different datasets that measured the electrophysiological responses to the exposure of an odorant [25, 26]. The experimental procedure used several different concentrations of odorants. 10^{-2} volume per volume (v/v) dilution with either H_2O, paraffin oil, or mineral oil is used, except for geosmin which used 10^{-3} dilution. We explicitly stated the dilution level whenever we used a different odorant concentration. When there was an overlap between the datasets, we chose the more recent recordings from Seki *et al.* [26] as the representative value. The connection between ORs and the respective glomerulus is identified through a literature search [73–76]. The dataset by Hallem *et al.* contained a recording corresponding to Or33b. Homotype DM3 coexpresses Or33b and Or47b, while homotype DM5 coexpresses Or33b and Or85a [74], which we incorporated manually by adjusting the odor response profile. When testing naturalistic inputs, we only chose odorants that elicited a strong response (≥ 40) to at least a single recorded homotype. This criterion resulted in odor response profiles for 39 homotypes to 96 odorants (+baseline firing rate). A keen reader might notice that neither of the datasets was comprehensive enough to cover all 58 homotypes available in the hemibrain dataset. For any missing measurements, we assumed no response and assigned zero values.

The functional group information was taken from Hallem *et al.* [25], while the odor valence information was acquired from the literature [6, 31, 32, 44, 46].

Algorithm of Compressed Sensing

Since the original inception by a group of mathematicians and engineers [77, 78], compressed sensing has garnered substantial interest due to the apparent overcoming of the Shannon-Nyquist sampling theorem [79, 80]. For a compressible signal that can be mapped to a sparse representation, CS aims to recover the original signal by reconstructing the sparsest solution of an under-determined linear

equation.

The novelty of CS is the utilization of the convex l_1 norm minimization to replace the non-convex optimization problem associated with l_0 norm minimization. Due to this property, many applications of CS have been proposed [81–83]. The first set of applications focuses on minimizing the quantity of measured input when the measurement process is expensive. Another set of applications recognizes the implementation of sparse representation for a high-dimensional signal. The primary example would be various attempts at analyzing the signal processing inside a brain. The brain has been suggested to benefit from utilizing CS as a means to reduce the bandwidth and storage necessary for neural processes, including vision [84], olfaction [10, 17, 20–23], and others [85, 86], thus inferring high-dimensional signals from low dimensional sensory input [56, 87].

Consider a linear algebra problem solving the equation

$$\mathbf{b} = \mathcal{A}\mathbf{x}, \tag{1}$$

where $\mathbf{x}$ is an unknown n-dimensional vector, $\mathbf{b}$ is a known m-dimensional vector, and $\mathcal{A}$ is a known $m \times n$ matrix. If our goal is to find $\mathbf{x}$ that satisfies the equation for given $\mathbf{b}$ and $\mathcal{A}$, the problem is trivial when the dimension of $\mathbf{b}$ is larger than the dimension of $\mathbf{x}$ ($m \geq n$). However, when $m < n$, this is an under-determined (over-parameterized) problem, allowing infinitely many solutions for $\mathbf{x}$. If our interest is to find $\hat{\mathbf{x}}$, the sparsest solution of $\mathbf{x}$ subject to Eq. 1 such that the vector $\mathbf{x}$ has the smallest number of nonzero values, our problem becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \mathbf{b} = \mathcal{A}\mathbf{x}, \tag{2}$$

where $\|\cdot\|_0$ is the cardinality or the l_0 pseudo-norm that counts the nonzero elements in a vector such that $\|\mathbf{x}\|_0 = K$ is the sparsity of the vector $\mathbf{x}$. However, minimizing the cardinality of x under a constraint corresponds to a non-convex optimization problem, which is computationally demanding. CS overcomes this computational barrier by adapting Eq. 2 with l_1 norm minimization instead, which algorithmically simplifies Eq. 2 to a convex optimization problem. That is, CS adapts Eq. 2 into

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{ subject to } \ \mathbf{b} = \mathcal{A}\mathbf{x}, \tag{3}$$

where $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$. If the following two conditions are met, solving Equation 3 is identical to solving Equation 2 and can recover the sparsest solution $\hat{\mathbf{x}}$ with a high probability [77, 80]. First, as the general rule of thumb, the dimension of $\mathbf{b}$ should satisfy a condition of $m > m_c$ with $m_c \approx K \log(n/K)$. Secondly, the sensing matrix $\mathcal{A}$ must be incoherent, such that the columns in sensing matrix $\mathcal{A}$ must be uncorrelated to each other, producing small inner products (see Supplementary Information).

In this study, we constructed the sensing matrix $\mathcal{A}$ by taking a product of the normalized synaptic connectivity matrices $\mathcal{C}^{\mathrm{PN-KC}}$ and $\mathcal{C}^{\mathrm{KC-MBON}}$. That is, each column of $\mathcal{C}^{\mathrm{PN-KC}}$ is normalized and each row of $\mathcal{C}^{\mathrm{KC-MBON}}$ is normalized. The l_1 norm minimization was performed to generate the solution $\hat{\mathbf{x}}^{\alpha}$ for a given MBON activity profile $\mathbf{b}^{meas,\alpha}$ simulated from uPN expression to a specific odorant. We used Sequential Least SQuares Programming (SLSQP) optimizer implemented in the Scipy package constrained to $\mathbf{b} = \mathcal{A}\mathbf{x}$ with unconstrained variable boundaries.

When assessing the effect of running the l_1 norm minimization using partial MBON, the vector $\mathbf{b}^{meas,\alpha}$ was randomly sampled to create a subset with the dimension $m = N_{\mathrm{MBON}}$. The sampling was performed ten times to create ten different subsets of MBON, and l_1 norm minimization and odor identification were attempted independently on each subset. We consider that the odorant is successfully identified or perceived without ambiguity if any subsets lead to the correct recovery.

Assessing the odor identifiability

The specificity (identifiability or differentiability) of an odor α against other odors is assessed by evaluating the similarity between (i) the *measurement* and (ii) the *calculation* of the MBON response profile. (i) The measured MBON response profile, denoted by $\mathbf{b}^{meas,\alpha}$, is acquired using the electrophysiological recording of uPN activity profile $\mathbf{x}^{meas,\alpha}$ and the sensing matrix $\mathcal{A}$ via the relation $\mathbf{b}^{meas,\alpha} = \mathcal{A}\mathbf{x}^{meas,\alpha}$. (ii) A MBON response profile of the odor α can be calculated from $\mathbf{b}^{\alpha} = \mathcal{A}\hat{\mathbf{x}}^{\alpha}$ using the solution of the CS algorithm, i.e., $\hat{\mathbf{x}}^{\alpha}$ from Eq.3.

To assess the differentiability of the odor α against other odors $\{\beta\}$, we adapted the sparse representation for classification (SRC) [80, 88] and considered a function $\delta_{\beta}(\hat{\mathbf{x}}^{\alpha})$ that filters the uPN activity profile calculated for the odor α against the uPN activity profile measured for the odor β. Specifically, the filtering function $\delta_{\beta}(\hat{\mathbf{x}}^{\alpha})$ for the i-th component of the uPN activity profile is defined as

$$[\delta_{\beta}(\hat{\mathbf{x}}^{\alpha})]_i = \begin{cases} \hat{x}_i^{\alpha} & x_i^{meas,\beta} \neq 0 \\ 0 & x_i^{meas,\beta} = 0. \end{cases} \tag{4}$$

Note that the self-filtering yields the original profile, i.e., $\delta_{\alpha}(\hat{\mathbf{x}}^{\alpha}) = \hat{\mathbf{x}}^{\alpha}$. Along with Eq.4, we defined $r_{\alpha|\beta}$, the (relative) residual between the measurement ($\mathbf{b}^{meas,\alpha}$) and the calculation ($\mathcal{A}\delta_{\beta}(\hat{\mathbf{x}}^{\alpha})$) over all test odors $\{\beta\}$ including $\beta = \alpha$, which enables us to assess the specificity of the odor α against all other odors $\{\beta\}$:

$$r_{\alpha|\beta} = \frac{\|\mathbf{b}^{meas,\alpha} - \mathcal{A}\delta_{\beta}(\hat{\mathbf{x}}^{\alpha})\|_2}{\|\mathbf{b}^{meas,\alpha}\|_2}, \tag{5}$$

where $\|\mathbf{x}\|_2 \equiv \sqrt{\sum_{i=1}^{n} x_i^2}$.

Ideally, $\mathcal{A}\delta_{\alpha}(\hat{\mathbf{x}}^{\alpha})$ must be identical to $\mathbf{b}^{meas,\alpha}$, i.e., $r_{\alpha|\alpha} = 0$, but in practice, we find $r_{\alpha|\alpha} \gtrsim 0$. However, if $r_{\alpha|\alpha} \ll r_{\alpha|\beta}$ for all $\beta \neq \alpha$, then the odor represented by α can be considered specific and easily discernible from other odors. In contrast, when there are many odors, say β_1, β_2, ..., satisfying $r_{\alpha|\alpha} \approx r_{\alpha|\beta_1} \approx r_{\alpha|\beta_2} \approx \cdots \approx r_{\alpha|\beta_k}$, it means that the odor space (perceptual space) of α is not well isolated from that of β, suggesting ambiguous sensing of odor α. In this case, discerning the odor α from other odors β_1, β_2, ..., β_k is challenging for *Drosophila*.

To assess the spectrum of the residuals for the odor α, we calculated the Z-score of $r_{\alpha|\alpha}$ defined as

$$Z = \frac{r_{\alpha|\alpha} - \mu_{r_{\alpha|\beta}}}{\sigma_{r_{\alpha|\beta}}}, \tag{6}$$

where $\mu_{r_{\alpha|\beta}} = \frac{1}{N_\beta}\sum_\beta r_{\alpha|\beta}$ and $\sigma_{r_{\alpha|\beta}} = \frac{1}{N_\beta - 1}\sum_\beta (r_{\alpha|\beta} - \mu_{r_{\alpha|\beta}})^2$. Therefore, a Z-score with $Z \ll 0$ implies that the value of the self-residual ($r_{\alpha|\alpha}$) is statistically well separated from the cross-residuals ($r_{\alpha|\beta}$), and the odor α is easily discernible from other odors.

Random odor mixtures

The odor mixtures are generated through a random selection of odorants. In this study, we generated 500 random odorant mixtures with a given size, N_{od}. Instead of residuals and Z-scores, we assessed the capacity of the CS framework through the cosine distances between $\mathbf{x}^{meas}$ and $\hat{\mathbf{x}}$, defined by

$$d_{\cos} = 1 - \frac{\mathbf{x}^{meas} \cdot \hat{\mathbf{x}}}{||\mathbf{x}^{meas}||_2 ||\hat{\mathbf{x}}||_2}. \tag{7}$$

When $d_{\cos} \leq 0.05$, the vector $\hat{\mathbf{x}}$ was considered sufficiently close to $\mathbf{x}^{meas}$.

DATA AVAILABILITY

All generated data used for the study are available at https://github.com/kirichoi/DrosophilaPerception.

CODE AVAILABILITY

The Python scripts used for the study are available at https://github.com/kirichoi/DrosophilaPerception.

ACKNOWLEDGMENTS

This study was supported by KIAS Individual Grants CG077002 (K.C.), CG076002 (W.K.K.), and CG035003 (C.H.). We thank the Center for Advanced Computation in KIAS for providing the computing resources.

AUTHOR CONTRIBUTIONS

K.C. and C.H. designed the project. K.C., W.K.K., and C.H. performed research, analyzed the data, and wrote the paper.

COMPETING INTERESTS

The authors declare that they have no competing interests.

SUPPLEMENTARY INFORMATION

Supplementary Information includes the text and twelve figures. All items are available at the end of this manuscript in the following order.

- Analysis of MBON response profiles using Euclidean distances
- Odor identification using partial information from MBONs
- Conditions for solving the inverse problem through compressed sensing
- Supplementary figures:
  Figure S1. PN-KC connectivity matrix.
  Figure S2. KC-MBON connectivity matrix.
  Figure S3. The MBON response profiles of all tested single odorants.
  Figure S4. Criteria required for successful signal recovery using the CS framework.
  Figure S5. The Z-scores of the natural mixtures at 10^{-2} dilution.
  Figure S6. The MBON response profiles of fruit extracts at 10^{-2} dilution.
  Figure S7. The sparsity K for select odorants at varying dilutions.
  Figure S8. The residuals $r_{\alpha|\beta}$ for undiluted fruit extracts.
  Figure S9. The MBON response profiles of single odorants at various dilution levels.
  Figure S10. Effect of noise on the success rate of odor identification.
  Figure S11. Analysis of MBON response profiles using the scaled Euclidean distances.
  Figure S12. Testing the odor identifiability using a random subset of MBONs.

References

[1] Bushdid, C., Magnasco, M. O., Vosshall, L. B. & Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).

[2] Gerkin, R. C. & Castro, J. B. The number of olfactory stimuli that humans can discriminate is still unknown. *Elife* **4**, e08127 (2015).

[3] Meister, M. On the dimensionality of odor space. *Elife* **4**, e07865 (2015).

[4] Owald, D. & Waddell, S. Olfactory learning skews mushroom body output pathways to steer behavioral choice in drosophila. *Curr. Opin. Neurobiol.* **35**, 178–184 (2015).

[5] Scheffer, L. K. *et al.* A connectome and analysis of the adult *Drosophila* central brain. *Elife* **9**, e57443 (2020).

[6] Bates, A. S. *et al.* Complete connectomic reconstruction of olfactory projection neurons in the fly brain. *Curr. Biol.* **30**, 3183–3199 (2020).

[7] Winding, M. *et al.* The connectome of an insect brain. *Science* **379**, eadd9330 (2023).

[8] Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the drosophila olfactory system. *Curr. Biol.* **15**, 1535–1547 (2005).

[9] Caron, S. J., Ruta, V., Abbott, L. F. & Axel, R. Random convergence of olfactory inputs in the Drosophila mushroom body. *Nature* **497**, 113–117 (2013).

[10] Stevens, C. F. What the fly's nose tells the fly's brain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9460–9465 (2015).

[11] Jefferis, G. S. *et al.* Comprehensive maps of drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128**, 1187–1203 (2007).

[12] Ruta, V. *et al.* A dimorphic pheromone circuit in Drosophila from sensory input to descending output. *Nature* **468**, 686–690 (2010).

[13] Liang, L. *et al.* Gabaergic projection neurons route selective olfactory inputs to specific higher-order neurons. *Neuron* **79**, 917–931 (2013).

[14] Choi, K., Kim, W. K. & Hyeon, C. Olfactory responses of $Drosophila$ are encoded in the organization of projection neurons. *Elife* **11**, e77748 (2022).

[15] Vosshall, L. B., Wong, A. M. & Axel, R. An olfactory sensory map in the fly brain. *Cell* **102**, 147–159 (2000).

[16] Wilson, C. D., Serrano, G. O., Koulakov, A. A. & Rinberg, D. A primacy code for odor identity. *Nat. Commun. ns* **8**, 1477 (2017).

[17] Kadakia, N. & Emonet, T. Front-end weber-fechner gain control enhances the fidelity of combinatorial odor coding. *Elife* **8**, e45293 (2019).

[18] Si, G. *et al.* Structured odorant response patterns across a complete olfactory receptor neuron population. *Neuron* **101**, 950–962 (2019).

[19] Dasgupta, S., Stevens, C. F. & Navlakha, S. A neural algorithm for a fundamental computing problem. *Science* **358**, 793–796 (2017).

[20] Tootoonian, S. & Lengyel, M. A dual algorithm for olfactory computation in the locust brain. *Advances in neural information processing systems* **27** (2014).

[21] Zhang, Y. & Sharpee, T. O. A robust feedforward model of the olfactory system. *PLoS Comp. Biol.* **12**, e1004850 (2016).

[22] Qin, S., Li, Q., Tang, C. & Tu, Y. Optimal compressed sensing strategies for an array of nonlinear olfactory receptor neurons with and without spontaneous activity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20286–20295 (2019).

[23] Krishnamurthy, K., Hermundstad, A. M., Mora, T., Walczak, A. M. & Balasubramanian, V. Disorder and the neural representation of complex odors. *Front. Comp. Neurosci.* **16** (2022).

[24] Dulac, C. Sparse encoding of natural scents. *Neuron* **50**, 816–818 (2006).

[25] Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).

[26] Seki, Y. *et al.* Olfactory coding from the periphery to higher brain centers in the $Drosophila$ brain. *BMC biology* **15**, 1–20 (2017).

[27] Shen, K., Tootoonian, S. & Laurent, G. Encoding of mixtures in a simple olfactory system. *Neuron* **80**, 1246–1262 (2013).

[28] De Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the Drosophila antenna. *Neuron* **30**, 537–552 (2001).

[29] Silbering, A. F. *et al.* Complementary function and integrated wiring of the evolutionarily distinct drosophila olfactory subsystems. *J. Neurosci.* **31**, 13357–13375 (2011).

[30] Fişek, M. & Wilson, R. I. Stereotyped connectivity and computations in higher-order olfactory neurons. *Nat. Neurosci.* **17**, 280–288 (2014).

[31] Galizia, C. G. & Sachse, S. Odor coding in insects. *The neurobiology of olfaction* **35**, 70 (2010).

[32] Badel, L., Ohta, K., Tsuchimoto, Y. & Kazama, H. Decoding of context-dependent olfactory behavior in Drosophila. *Neuron* **91**, 155–167 (2016).

[33] Dolan, M.-J. *et al.* Neurogenetic dissection of the drosophila lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. *Elife* **8**, e43079 (2019).

[34] Schlegel, P. *et al.* Information flow, cell types and stereotypy in a full olfactory connectome. *Elife* **10**, e66018 (2021).

[35] Laing, D. G. & Francis, G. The capacity of humans to identify odors in mixtures. *Physiology & behavior* **46**, 809–814 (1989).

[36] Laing, D. & Livermore, B. Perceptual analysis of complex chemical signals by humans. *Chemical Signals in Vertebrates 6* 587–593 (1992).

[37] Poupon, D., Fernandez, P., Archambault Boisvert, S., Migneault-Bouchard, C. & Frasnelli, J. Can the identification of odorants within a mixture be trained? *Chemical Senses* **43**, 721–726 (2018).

[38] Weiss, T. *et al.* Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19959–19964 (2012).

[39] Tabor, R., Yaksi, E., Weislogel, J.-M. & Friedrich, R. W. Processing of odor mixtures in the zebrafish olfactory bulb. *J. Neurosci.* **24**, 6611–6620 (2004).

[40] Joerges, J., Küttner, A., Galizia, C. G. & Menzel, R. Representations of odours and odour mixtures visualized in the honeybee brain. *Nature* **387**, 285–288 (1997).

[41] Deisig, N., Giurfa, M. & Sandoz, J. C. Antennal lobe processing increases separability of odor mixture representations in the honeybee. *J. Neurophys.* **103**, 2185–2194 (2010).

[42] Thomas-Danguin, T. *et al.* The perception of odor objects in everyday life: a review on the processing of odor mixtures. *Front. Psychol.* **5**, 504 (2014).

[43] Wilson, R. I. Early olfactory processing in $Drosophila$: mechanisms and principles. *Annu. Rev. Neurosci.* **36**, 217–241 (2013).

[44] Hallem, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the drosophila antenna. *Cell* **117**, 965–979 (2004).

[45] Knaden, M., Strutz, A., Ahsan, J., Sachse, S. & Hansson, B. S. Spatial representation of odorant valence in an insect brain. *Cell reports* **1**, 392–399 (2012).

[46] Mansourian, S. & Stensmyr, M. C. The chemical ecology of the fly. *Curr. Opin. Neurobiol.* **34**, 95–102 (2015).

[47] Chakraborty, S. D., Chang, H., Hansson, B. S. & Sachse, S. Higher-order olfactory neurons in the lateral horn support odor valence and odor identity coding in drosophila. *Elife* **11**, e74637 (2022).

[48] Ai, M. *et al.* Acid sensing by the drosophila olfactory system. *Nature* **468**, 691–695 (2010).

[49] Yoshida, K. *et al.* Odour concentration-dependent olfactory preference change in C. elegans. *Nat. Commun.* **3**, 739 (2012).

[50] Bontempi, C., Jacquot, L. & Brand, G. A study on the relationship between odor hedonic ratings and individual odor detection threshold. *Sci. Rep.* **12**, 18482 (2022).

[51] Kazama, H. & Wilson, R. I. Homeostatic matching and nonlinear amplification at identified central synapses. *Neuron* **58**, 401–413 (2008).

[52] Perisse, E., Burke, C., Huetteroth, W. & Waddell, S. Shocking revelations and saccharin sweetness in the study of drosophila olfactory memory. *Current Biology* **23**, R752–R763 (2013).

[53] MacWilliam, D., Kowalewski, J., Kumar, A., Pontrello, C. & Ray, A. Signaling mode of the broad-spectrum conserved CO2 receptor is one of the important determinants of odor valence in Drosophila. *Neuron* **97**, 1153–1167 (2018).

[54] Pehlevan, C., Genkin, A. & Chklovskii, D. B. A clustering neural network model of insect olfaction. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 593–600 (IEEE, 2017).

[55] Koulakov, A. A. & Rinberg, D. Sparse incomplete representations: a potential role of olfactory granule cells. *Neuron* **72**, 124–136 (2011).

[56] Ganguli, S. & Sompolinsky, H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* **35**, 485–508 (2012).

[57] Ghosal, P., Mahankali, S. & Sun, Y. Randomly initialized one-layer neural networks make data linearly separable. *arXiv preprint arXiv:2205.11716* (2022).

[58] Silbering, A. F. & Galizia, C. G. Processing of odor mixtures in the Drosophila antennal lobe reveals both global inhibition and glomerulus-specific interactions. *J. Neurosci.* **27**, 11966–11977 (2007).

[59] Linster, C. & Cleland, T. A. Configurational and elemental odor mixture perception can arise from local inhibition. *J. Comp. Neurosci.* **16**, 39–47 (2004).

[60] Olsen, S. R., Bhandawat, V. & Wilson, R. I. Divisive normalization in olfactory population codes. *Neuron* **66**, 287–299 (2010).

[61] Berck, M. E. *et al.* The wiring diagram of a glomerular olfactory system. *Elife* **5**, e14859 (2016).

[62] Mohamed, A. A. *et al.* Odor mixtures of opposing valence unveil interglomerular crosstalk in the Drosophila antennal lobe. *Nat. Commun.* **10**, 1201 (2019).

[63] Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolinsky, H. & Abbott, L. Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153–1164 (2017).

[64] Amin, H., Apostolopoulou, A. A., Suárez-Grimalt, R., Vrontou, E. & Lin, A. C. Localized inhibition in the $Drosophila$ mushroom body. *Elife* **9**, e56954 (2020).

[65] Hong, E. J. & Wilson, R. I. Simultaneous encoding of odors by channels with diverse sensitivity to inhibition. *Neuron* **85**, 573–589 (2015).

[66] Cury, K. M. & Uchida, N. Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* **68**, 570–585 (2010).

[67] Parabucki, A. *et al.* Odor concentration change coding in the olfactory bulb. *Eneuro* **6** (2019).

[68] Dalton, P. Psychophysical and behavioral characteristics of olfactory adaptation. *Chemical senses* **25**, 487–492 (2000).

[69] Liu, T. X., Davoudian, P. A., Lizbinski, K. M. & Jeanne, J. M. Connectomic features underlying diverse synaptic connection strengths and subcellular computation. *Current Biology* **32**, 559–569 (2022).

[70] Kim, S. H., Woo, J., Choi, K., Choi, M. & Han, K. Neural information processing and computations of two-input synapses. *Neural Computation* **34**, 2102–2131 (2022).

[71] Quinn, W. G., Harris, W. A. & Benzer, S. Conditioned behavior in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 708–712 (1974).

[72] Tully, T. & Quinn, W. G. Classical conditioning and retention in normal and mutant drosophila melanogaster. *J. Comp. Physiol. A* **157**, 263–277 (1985).

[73] Ebrahim, S. A. *et al.* Drosophila avoids parasitoids by sensing their semiochemicals via a dedicated olfactory circuit. *PLoS biology* **13**, e1002318 (2015).

[74] Grabe, V. *et al.* Elucidating the neuronal architecture of olfactory glomeruli in the drosophila antennal lobe. *Cell reports* **16**, 3401–3413 (2016).

[75] Coates, K. E. *et al.* Identified serotonergic modulatory neurons have heterogeneous synaptic connectivity within the olfactory system of Drosophila. *J. Neurosci.* **37**, 7318–7331 (2017).

[76] Hayashi, T. T. *et al.* Mushroom body input connections form independently of sensory activity in drosophila melanogaster. *Current Biology* **32**, 4000–4012 (2022).

[77] Candes, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* **59**, 1207–1223 (2006).

[78] Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).

[79] Shannon, C. E. A Mathematical Theory of Communication. *Bell Sys. Tech. J.* **27**, 379–423 (1948).

[80] Brunton, S. L. & Kutz, J. N. *Data-driven science and engineering: Machine learning, dynamical systems, and control* (Cambridge University Press, 2022).

[81] Lustig, M., Donoho, D. & Pauly, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **58**, 1182–1195 (2007).

[82] Leary, R., Saghi, Z., Midgley, P. A. & Holland, D. J. Compressed sensing electron tomography. *Ultramicroscopy* **131**, 70–91 (2013).

[83] Lee, J. B. *et al.* A compressed sensing framework for efficient dissection of neural circuits. *Nature methods* **16**, 126–133 (2019).

[84] Barranca, V. J., Kovačič, G., Zhou, D. & Cai, D. Efficient image processing via compressive sensing of integrate-and-fire neuronal network dynamics. *Neurocomputing* **171**, 1313–1322 (2016).

[85] Petrantonakis, P. C. & Poirazi, P. A compressed sensing perspective of hippocampal function. *Front. Syst. Neurosci.* **8**, 141 (2014).

[86] Sun, B. & Zhao, W. Compressed sensing of extracellular neurophysiology signals: A review. *Front. Neurosci.* **15**, 682063 (2021).

[87] Barranca, V. J., Kovačič, G., Zhou, D. & Cai, D. Sparsity and compressed coding in sensory systems. *PLoS Comp. Biol.* **10**, e1003793 (2014).

[88] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. & Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 210–227 (2008).

[89] Candes, E. J. & Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005).

[90] Candès, E. J., Romberg, J. & Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory* **52**, 489–509 (2006).

[91] Candes, E. J., Eldar, Y. C., Needell, D. & Randall, P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis* **31**, 59–73 (2011).

[92] Stankovic, L., Mandic, D. P., Dakovic, M. & Kisil, I. Demystifying the coherence index in compressive sensing [lecture notes]. *IEEE Signal Process. Mag.* **37**, 152–162 (2020).

[93] Adcock, B., Hansen, A. C., Poon, C. & Roman, B. Breaking the coherence barrier: A new theory for compressed sensing. In *Forum of Mathematics, Sigma*, vol. 5, e4 (Cambridge University Press, 2017).

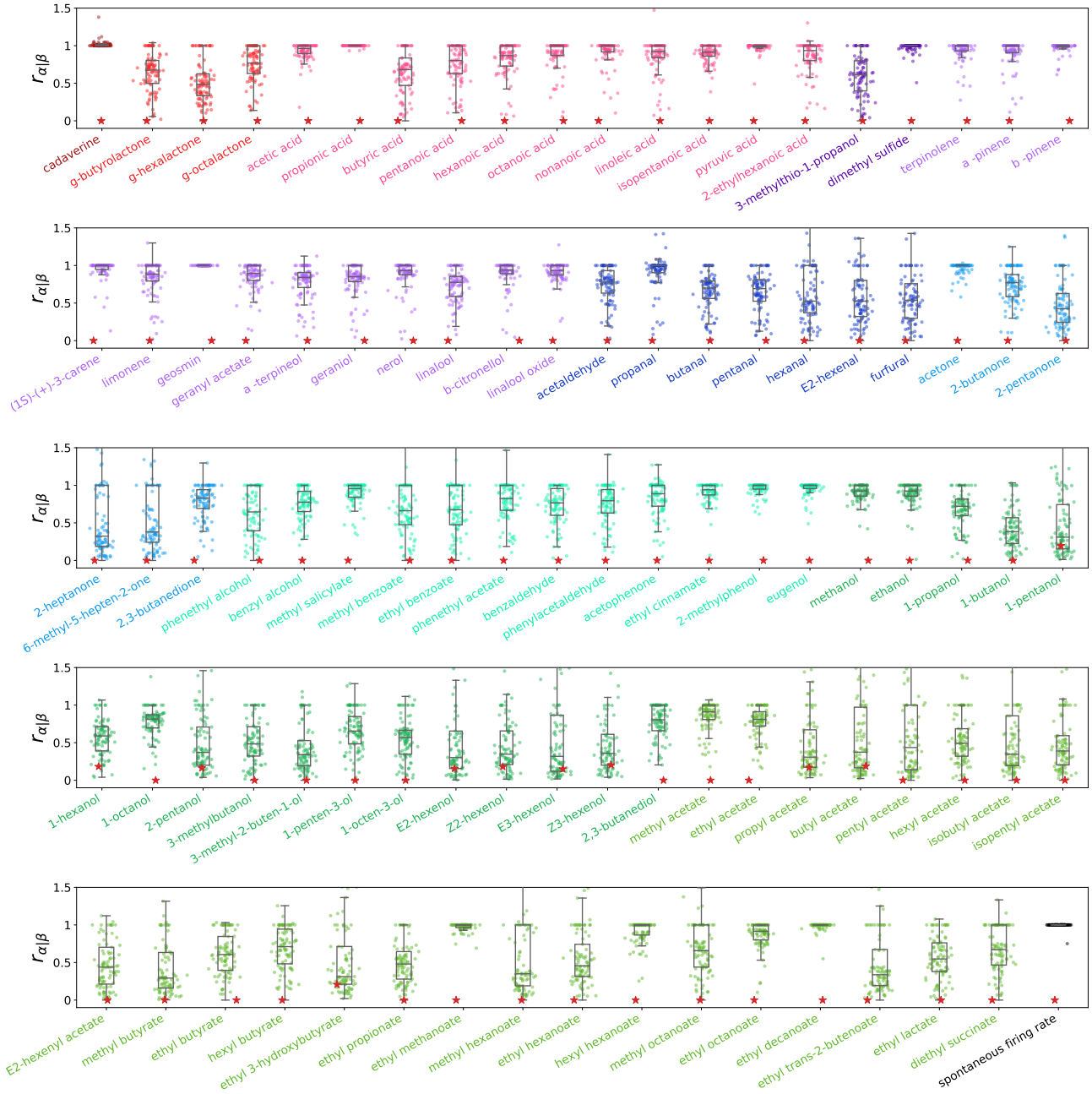
[94] Candes, E. J. & Plan, Y. A probabilistic and ripless theory of compressed sensing. *IEEE Trans. Inf. Theory* **57**, 7235–7254 (2011).

EXTENDED DATA

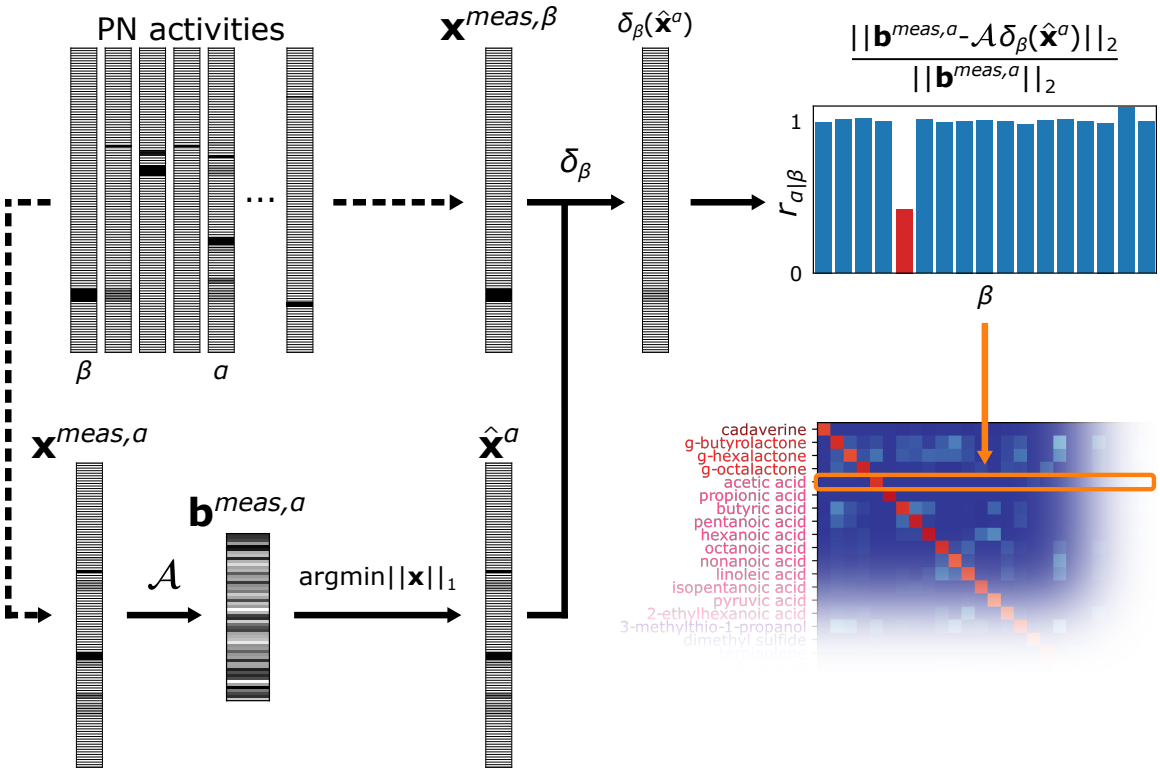
Extended Table 1: The list of odorants.

Odorant	Functional group	Odorant	Functional group
cadaverine‡	amines	phenethyl acetate	aromatics
γ-butyrolactone†	lactones	benzaldehyde*	aromatics
γ-hexalactone‡	lactones	phenylacetaldehyde†	aromatics
γ-octalactone	lactones	acetophenone‖	aromatic
acetic acid‖	acids	ethyl cinnamate‡	aromatics
propionic acid†	acids	2-methylphenol*	aromatics
butyric acid‡	acids	eugenol‡	aromatics
pentanoic acid‡	acids	methanol‡	alcohols
hexanoic acid‡	acids	ethanol‖	alcohols
octanoic acid‡	acids	1-propanol‡	alcohols
nonanoic acid	acids	1-butanol‡	alcohols
linoleic acid	acids	1-pentanol†	alcohols
isopentanoic acid‡	acids	1-hexanol	alcohols
pyruvic acid‖	acids	1-octanol*	alcohols
2-ethylhexanoic acid‡	acids	2-pentanol‡	alcohols
3-methylthio-1-propanol‡	surfur compounds	3-methylbutanol‡	alcohols
dimethyl sulfide	surfur compounds	3-methyl-2-buten-1-ol‡	alcohols
terpinolene	terpenes	1-penten-3-ol‡	alcohols
α-pinene‡	terpenes	1-octen-3-ol*	alcohols
β-pinene‡	terpenes	trans-2-hexen-1-ol	alcohols
(1S)-(+)-3-carene	terpenes	cis-2-hexen-1-ol‡	alcohols
limonene	terpenes	trans-3-hexen-1-ol	alcohols
geosmin*	terpenes	cis-3-hexen-1-ol	alcohols
geranyl acetate‖	terpenes	2,3-butanediol‡	alcohols
α-terpineol‡	terpenes	methyl acetate†	esters
geraniol	terpenes	ethyl acetate†	esters
nerol	terpenes	propyl acetate†	esters
linalool*	terpeness	butyl acetate	esters
β-citronellol	terpenes	pentyl acetate‡	esters
linalool oxide‡	terpenes	hexyl acetate‖	esters
acetaldehyde	aldehydes	isobutyl acetate	esters
propanal	aldehydes	isopentyl acetate	esters
butanal‡	aldehydes	trans-2-hexenyl acetate	esters
pentanal	aldehydes	methyl butyrate	esters
hexanal‡	aldehydes	ethyl butyrate‖	esters
trans-2-hexenal	aldehydes	hexyl butyrate†	esters
furfural‡	aldehydes	ethyl 3-hydroxybutyrate‖	esters
acetone‡	ketones	ethyl propionate‡	esters
2-butanone	ketones	ethyl methanoate‡	esters
2-pentanone‡	ketones	methyl hexanoate	esters
2-heptanone†	ketones	ethyl hexanoate	esters
6-methyl-5-hepten-2-one	ketones	hexyl hexanoate‡	esters
2,3-butanedione†	ketones	methyl octanoate	esters
phenethyl alcohol‡	aromatics	ethyl octanoate	esters
benzyl alcohol‡	aromatics	ethyl decanoate	esters
methyl salicylate*	aromatics	ethyl trans-2-butenoate	esters
methyl benzoate‡	aromatics	ethyl lactate	esters
ethyl benzoate‖	aromatics	diethyl succinate‡	esters

† Attractive w/o dilution.
‡ Reported to be attractive but could not be verified at high concentrations.
* Aversive w/o dilution.
‖ Odorants with conflicting valence information.



Extended Figure 1: The distributions of $r_{\alpha|\beta}$ for all tested single odorants. Red stars denote the self-residual $r_{\alpha|\alpha}$.



Extended Figure 2: The illustration of our computational pipeline. uPN activity profile is used to calculate the input MBON response profile $\mathbf{b}^{meas,\alpha}$, where α is an odorant. l_1 norm minimization is performed to determine the solution $\hat{\mathbf{x}}^\alpha$ from a set of underdetermined linear equations, which is compared with the input uPN activity profile $\delta_\beta(\hat{\mathbf{x}}^\alpha)$ filtered based on the expected uPN activity from another odorant β, $\mathbf{x}^{meas,\beta}$. Adapting the framework of sparse representation for classification (SRC) [80, 88], we calculate the residuals $r_{\alpha|\beta}$ for a given input odorant α and reference odorant β. The residuals for an odorant α against all other reference odorants $\{\beta\}$, $r_{\alpha|\beta}$, are calculated.

SUPPLEMENTARY INFORMATION

Analysis of MBON response profiles using Euclidean distances

In this study, we utilize CS to discuss the differentiability between the MBON response profiles of various odorants. In principle, one could perform a similar analysis by calculating the distances between MBON response profiles. We, however, find that the results from such an analysis on high dimensional vectors are often limited by the issues associated with the curse of dimensionality [56]. We find that the recovery of the uPN activity through the CS algorithm is better suited for odor differentiability than other standardized methods.

Consider a pairwise distance matrix between MBON response profiles. Supplementary Figure S11A shows the distribution of the scaled Euclidean distances between uPN activity profiles, $\hat{d}(\mathbf{x}^{meas,\alpha}, \mathbf{x}^{meas,\beta})$, where

$$\hat{d}(\mathbf{u}, \mathbf{v}) = d(\mathbf{u}, \mathbf{v}) / \max_{\mathbf{v}}(d(\mathbf{u}, \mathbf{v})), \tag{8}$$

with $d(\mathbf{u}, \mathbf{v}) = [\sum_{i=1}^{n}(u_i - v_i)^2]^{1/2}$ and $\max_{\mathbf{v}}(d(\mathbf{u}, \mathbf{v}))$ being the maximum among the distances from the vector $\mathbf{u}$ to other vectors $\mathbf{v}$. The clear gap between the self- ($\hat{d}(\mathbf{x}^{meas,\alpha}, \mathbf{x}^{meas,\alpha}) = 0$) and cross-distance ($\hat{d}(\mathbf{x}^{meas,\alpha}, \mathbf{x}^{meas,\beta}) > 0$) found in the uPN activity profiles is no longer observable in the MBON response profiles ($\hat{d}(\mathbf{b}^{meas,\alpha}, \mathbf{b}^{meas,\beta})$). This suggests that the Euclidean distance and its derivative measures are inadequate to characterize the highly integrated signal at MBONs (Supplementary Figure S11A, bottom). We observed a similar trend when we used other metrics such as cosine distance.

Incidentally, separation of the odor space by physiochemical features of odorants and OR/PN activities have previously been shown using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) [18, 25, 45, 47]. We performed PCA and t-SNE on a standardized matrix of MBON response profiles, in which each column of the matrix that represents MBON response profiles has zero mean and unit variance. PCA was performed using singular value decomposition (SVD), and t-SNE was performed using a perplexity of 15 after initializing with PCA. The result of PCA indicates that, unlike uPN activity profiles, many odorants are hard to differentiate using MBON response profiles (Supplementary Figure S11B), where the proportion of variance explained by each principal component quickly diminishes after the first principal component (Supplementary Figure S11B, bottom), suggestive of no clear separation in the odor representation. The lack of a clear separation in the component space is also found in the analysis using t-SNE (Supplementary Figure S11C).

Odor identification using partial information from MBONs

Our CS framework reconstructs the uPN activity from the response profile of 56 MBONs, where we assume that the MBON response profile reflects the information transmitted to the higher cortical area. Here, we consider a hypothetical situation where only partial information from the MBONs is available to the higher cortical area, perhaps due to some stereotyped connectivity, intrinsic stochasticity, and limited access to MBONs at a given timeframe in animal perception, and ask whether the information encoded by a subset of MBONs is still enough to accurately infer the uPN activity profile [10]. For each of the 86 odorants our CS framework correctly identified, we randomly sample the MBON response profile 10 times with a sampling size N_{MBON} and examine whether at least one of the samples led to a correct identification. We find that a large number of odors can still be identified by using only a handful of MBONs, with the success rate improving with the sampling size N_{MBON} (see Supplementary Figure S12A). We observe that the compressibility of information necessary to identify the odor varies with the functional group of the odorants. In particular, alcohols and esters are the first to fail as the size of N_{MBON} decreases (see Supplementary Figure S12B). Odors from other functional groups (e.g., acids and terpenes) retain their accuracy even if the size of N_{MBON} is significantly small.

Conditions for solving the inverse problem through compressed sensing

We performed a detailed analysis of the synaptic connectivity between the layers of olfactory neurons to check whether it conforms to the criteria for successful recovery of the sparsest solution $\hat{\mathbf{x}}$ through CS (see Methods for a detailed explanation). Unlike the previous applications of CS to the olfactory system [17, 22, 23], where the true dimension of the odorant space is unknown, the dimensions of input and output vectors are well-defined in our problem.

First, CS requires that the size of the input (m, the dimension of $\mathbf{b}$) be reasonably large compared to the dimension (n) and the sparsity (K) of the target vector to be recovered ($\mathbf{x}$). According to the synaptic connectivity provided by *Drosophila* hemibrain connectome [5], out of 162 uPNs, only 119 uPNs innervate MB calyx and connect to KCs [14]. These $n = 119$ uPNs that synapse with at least one of ~ 2000 KCs can be traced down to $m = 56$ MBONs based on the synaptic connectivity. We find that $m > m_c \approx K \log(n/K)$, the empirical criterion of CS for a successful recovery of the $\hat{\mathbf{x}}$ [89, 90], is satisfied for all values of K (Supplementary Figure S4A).

We also approximated the incoherence criterion of the sensing matrix $\mathcal{A}$ constructed from two connectivity matrices, $\mathcal{C}^{\mathrm{PN-KC}}$ and $\mathcal{C}^{\mathrm{KC-MBON}}$, between uPNs, KCs, and MBONS. In the original framework of CS, the incoherence of the sensing matrix $\mathcal{A}$ is defined through a value [91, 92]

$$\mu = \max_{j,k,\, j \neq k} \mathcal{M}_{j,k}, \tag{9}$$

where

$$\mathcal{M}_{j,k} = \frac{|\mathcal{A}_{,j} \cdot \mathcal{A}_{,k}|}{||\mathcal{A}_{,j}||_2 ||\mathcal{A}_{,k}||_2}. \tag{10}$$

Here, $\mathcal{A}_{,i}$ denotes the i-th column vector of $\mathcal{A}$. The value μ ranges from 0 to 1, with 0 indicating complete incoherence. Then, recon-

structing a K-sparse vector $\hat{\mathbf{x}}$ becomes highly probable under the following condition [92]:

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \tag{11}$$

From the sensing matrix $\mathcal{A} = (\mathcal{C}^{\mathrm{PN-KC}} \cdot \mathcal{C}^{\mathrm{KC-MBON}})^{\intercal}$ and Equation 9, we obtain $\mu \approx 1$, which is the theoretical maximum. Based on this result, the reconstruction of the vector $\hat{\mathbf{x}}$ through l_1 norm minimization seems infeasible. However, the requirement for incoherence defined in Eq. 9 is often considered too stringent to achieve, given that $\mu = 1$ is obtained even when a single pair of columns in the sensing matrix $\mathcal{A}$ are identical. As a result, several guidelines to relax the original criterion for incoherence have been suggested [91, 93], requiring that the sensing matrix be only *effectively incoherent*.

In practice, it has been suggested that the recovery of a K-sparse vector is achievable with a high probability from asymptotically small μ as the size of $\mathcal{A}$ increases [92, 93]. Brunton and Kutz have suggested a more relaxed condition [80]. When the sensing matrix $\mathcal{A}$ is a product of a measurement matrix $\mathcal{S}$ and a sparsifying basis Ψ, i.e., $\mathcal{A} = \mathcal{S}\Psi$, a small scalar product between the rows of $\mathcal{S}$ and the columns of Ψ is a good indication that l_1 norm minimization will recover $\hat{\mathbf{x}}$, which replaces the matrix $\mathcal{M}_{j,k}$ (Eq. 10) with

$$\mathcal{M}^{*}_{j,k} = \frac{|\mathcal{S}_{j,} \cdot \Psi_{,k}|}{||\mathcal{S}_{j,}||_2 ||\Psi_{,k}||_2}. \tag{12}$$

While our $\mathcal{C}^{\mathrm{PN-KC}}$ and $\mathcal{C}^{\mathrm{KC-MBON}}$ are not analogous to the measurement matrix and the sparsifying basis by definition, the suggested requirement is mathematically equivalent within the scope of successful l_1 norm minimization subject to $\mathbf{b} = \mathcal{A}\mathbf{x}$. When we adopt $\mathcal{M}^{*}_{j,k}$ to approximate the incoherence criterion, a large portion of $\mathcal{M}^{*}_{j,k}$ are zero or close to zero (Supplementary Figure S4B), and about 96% of the values are below 0.2, a theoretical limit which would allow CS to recover a target vector $\mathbf{x}$ with the sparsity of $K = 3$ (Eq. 11), which corresponds to an approximate average uPNs per glomerulus.

Restricted isometry property (RIP) is often cited as the rule of thumb when it comes to solving the inverse problem, but checking for RIP is often prohibitively expensive [92] and unnecessary in practical applications [93]. It has, indeed, been demonstrated that CS is possible even under weaker conditions [93, 94]. We find that the sensing matrix of the *Drosophila* olfactory system is effective in recovering $\hat{\mathbf{x}}$ for all practical purposes.

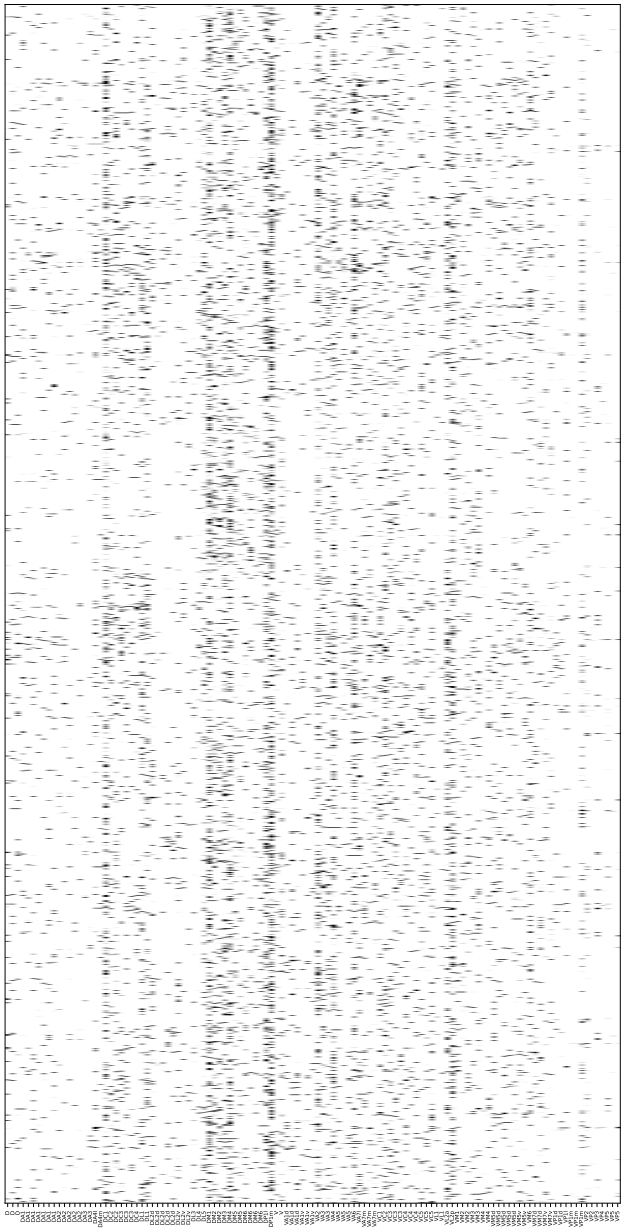


Figure S1: **The** 119×1776 **PN-KC connectivity matrix** ($\mathcal{C}^{\mathrm{PN-KC}}$). The values are normalized to the mean synaptic weight for better visibility.

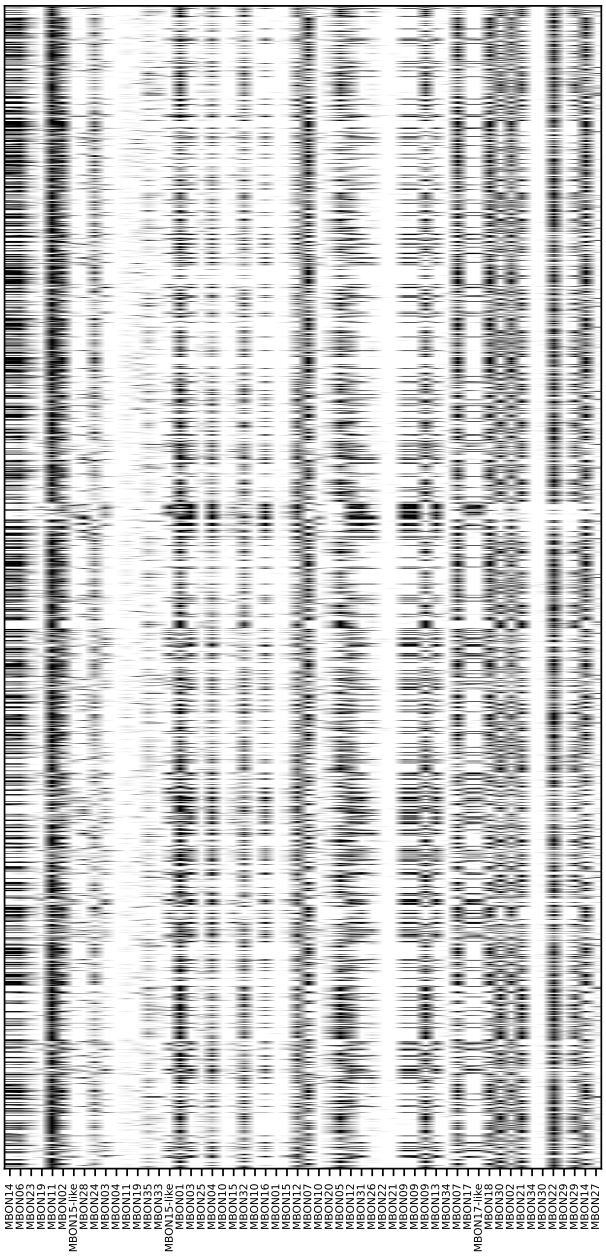


Figure S2: **The** 1776×56 **KC-MBON connectivity matrix** ($\mathcal{C}^{\mathrm{KC-MBON}}$). The values are normalized to the mean synaptic weight for better visibility.

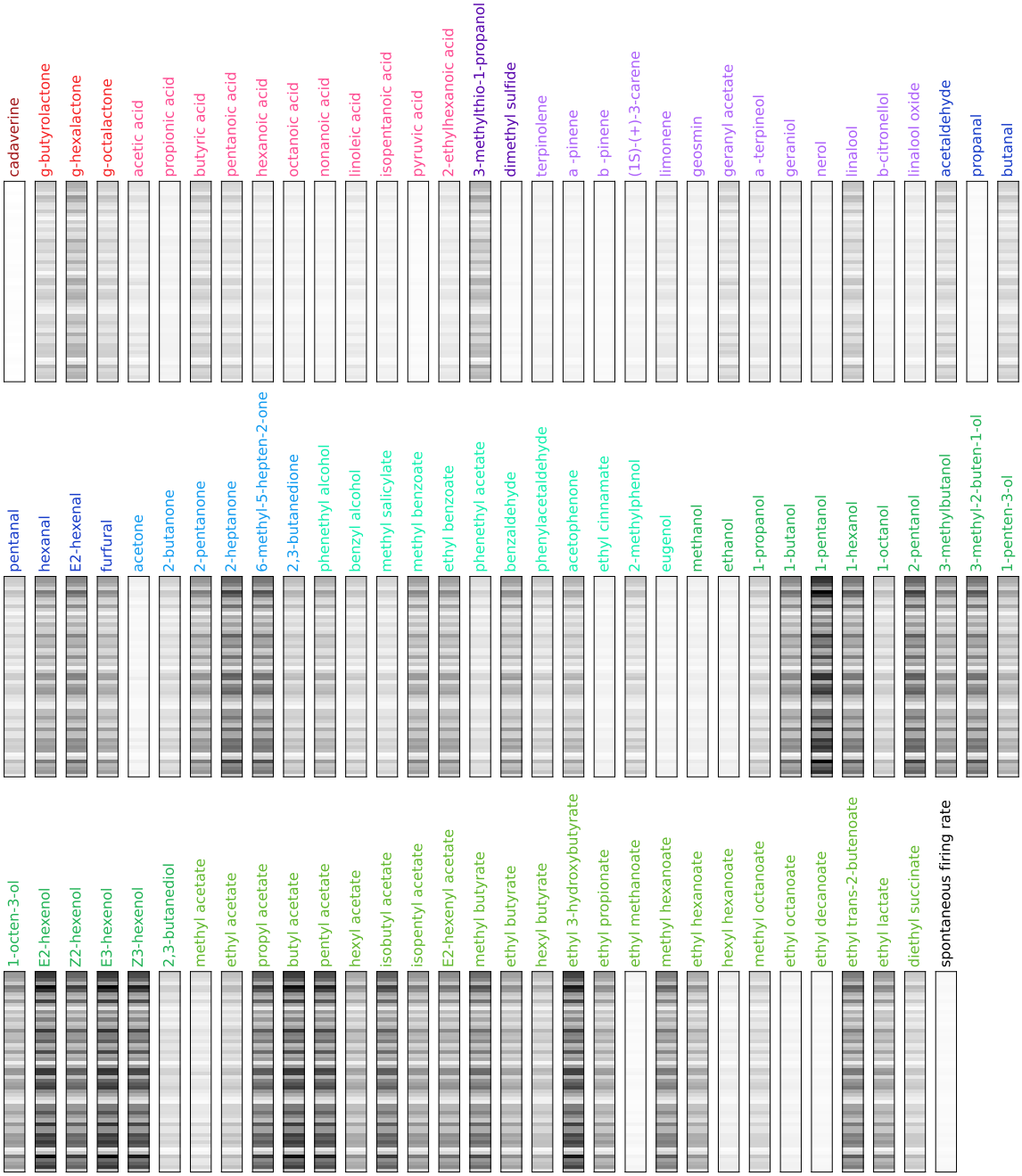


Figure S3: **The MBON response profiles of all tested single odorants.** The labels are color-coded based on the same odor categorization used by Hallem *et al.* [25] (Dark red: amines, red: lactones, pink: acids, purple: sulfur compounds, violet: terpenes, dark blue: aldehydes, blue: ketones, emerald: aromatics, green: alcohols, olive: esters)

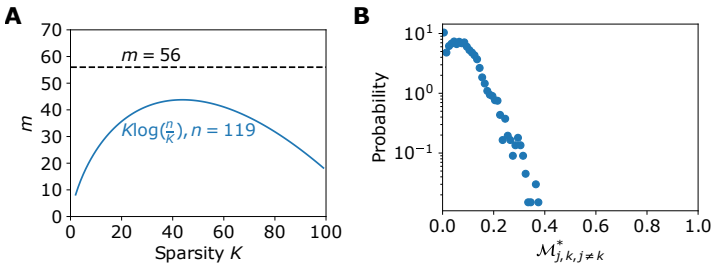


Figure S4: **Criteria required for successful signal recovery using the CS framework.** (A) A plot illustrating the gap between the dimension of MBON ($m = 56$) and the K-dependent curve $K \log(n/K)$ for $n = 119$. (B) The distribution of the pairwise correlation $\mathcal{M}^*_{j,k}$ between the connectivity matrices $\mathcal{C}^{\mathrm{PN-KC}}$ and $\mathcal{C}^{\mathrm{KC-MBON}}$.

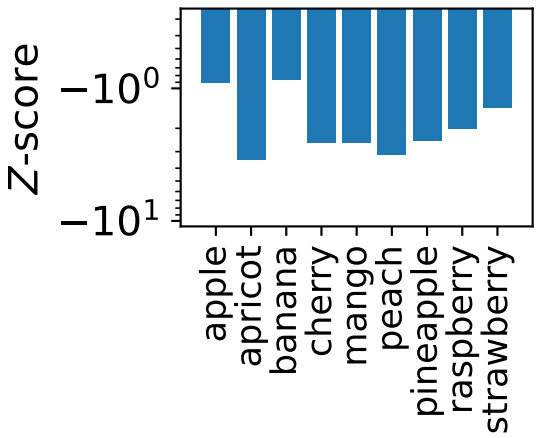


Figure S5: **The Z-scores of the natural mixtures at 10^{-2} dilution.**

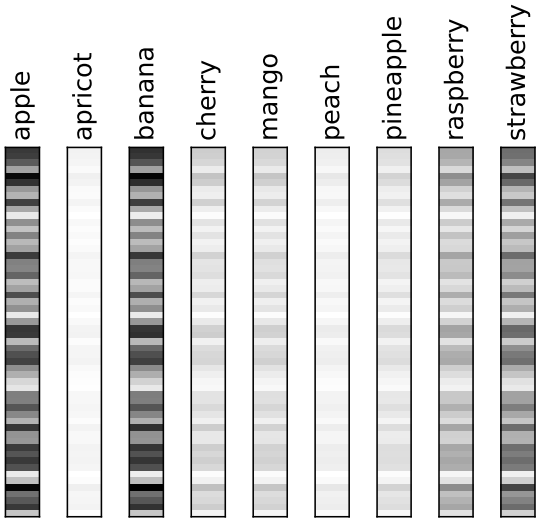


Figure S6: **The MBON response profiles of fruit extracts at 10^{-2} dilution.**

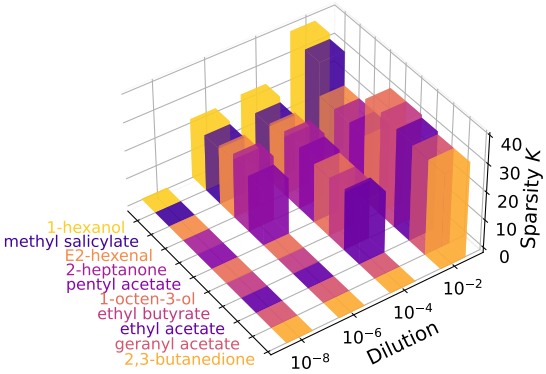


Figure S7: **The sparsity K for select odorants at varying dilutions.**

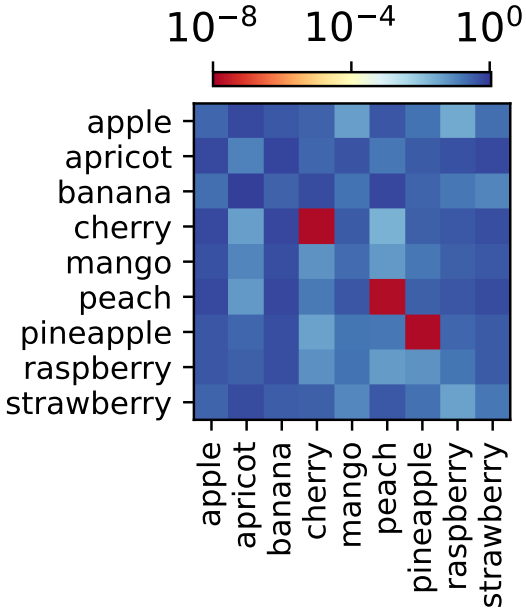


Figure S8: **The residuals $r_{\alpha|\beta}$ for undiluted fruit extracts.** Each row depicts the spectrum of the residuals of an odor input α with respect to other test odors β ($\beta =$ apple, apricot, ...) given on the horizontal axis.

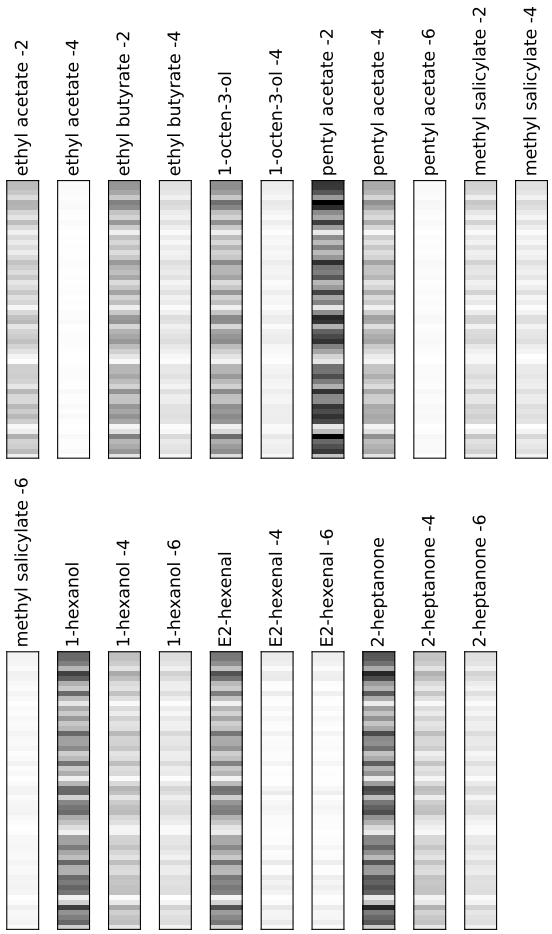


Figure S9: **The MBON response profiles of single odorants at various dilution levels.** -2, -4, and -6 signify the dilution levels of 10^{-2}, 10^{-4}, and 10^{-6}, respectively.

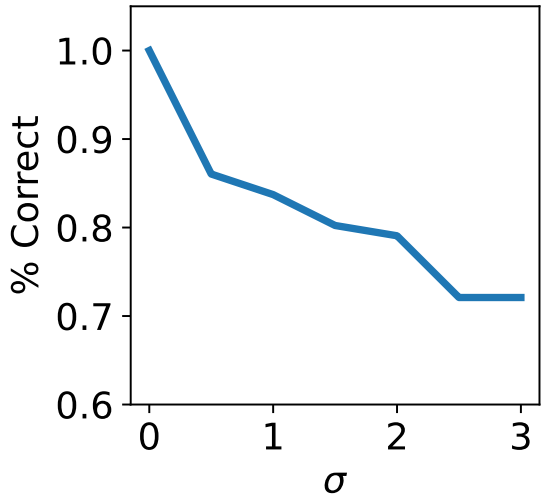


Figure S10: **Effect of noise on the success rate of odor identification.** σ denotes the strength of noise added to the uPN activity profile.

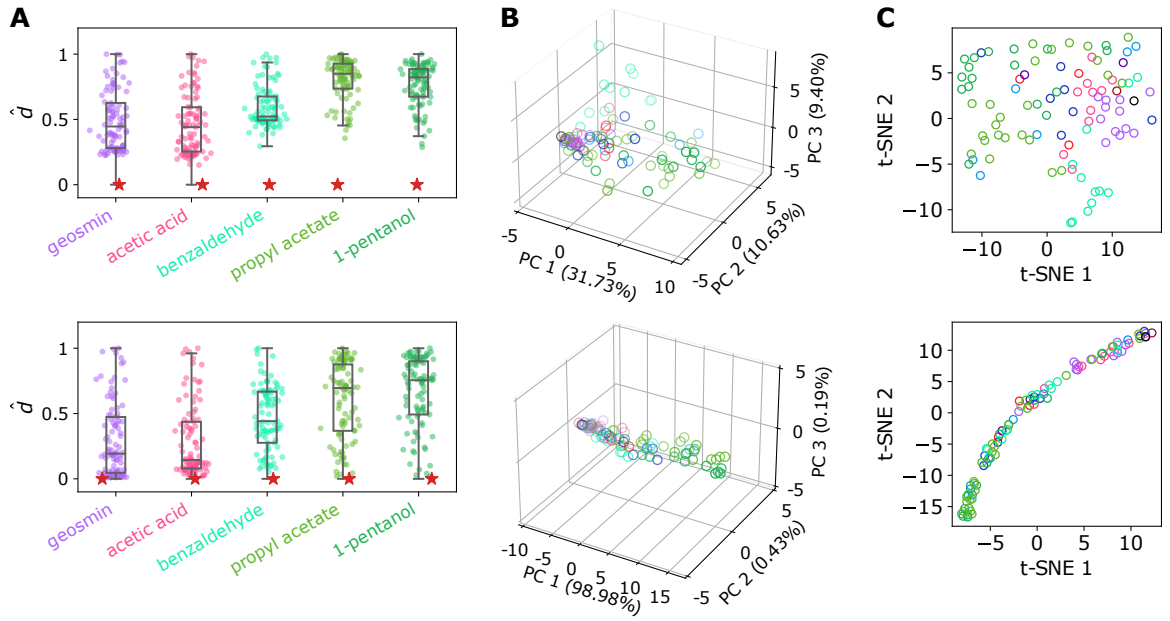


Figure S11: **Analysis of MBON response profiles using the scaled Euclidean distances.** (A) The distribution of scaled Euclidean distances between (top) uPN activity profiles and (bottom) MBON response profiles in response to selected odorants. The same color code as Figure 3 is used. (B) The output of PCA projecting (top) uPN activity profiles and (bottom) MBON response profiles to 3-dimensional space, along with the proportion of variance explained by each principal component. (C) The output of t-SNE projecting (top) uPN activity profiles and (bottom) MBON response profiles to 2-dimensional space.

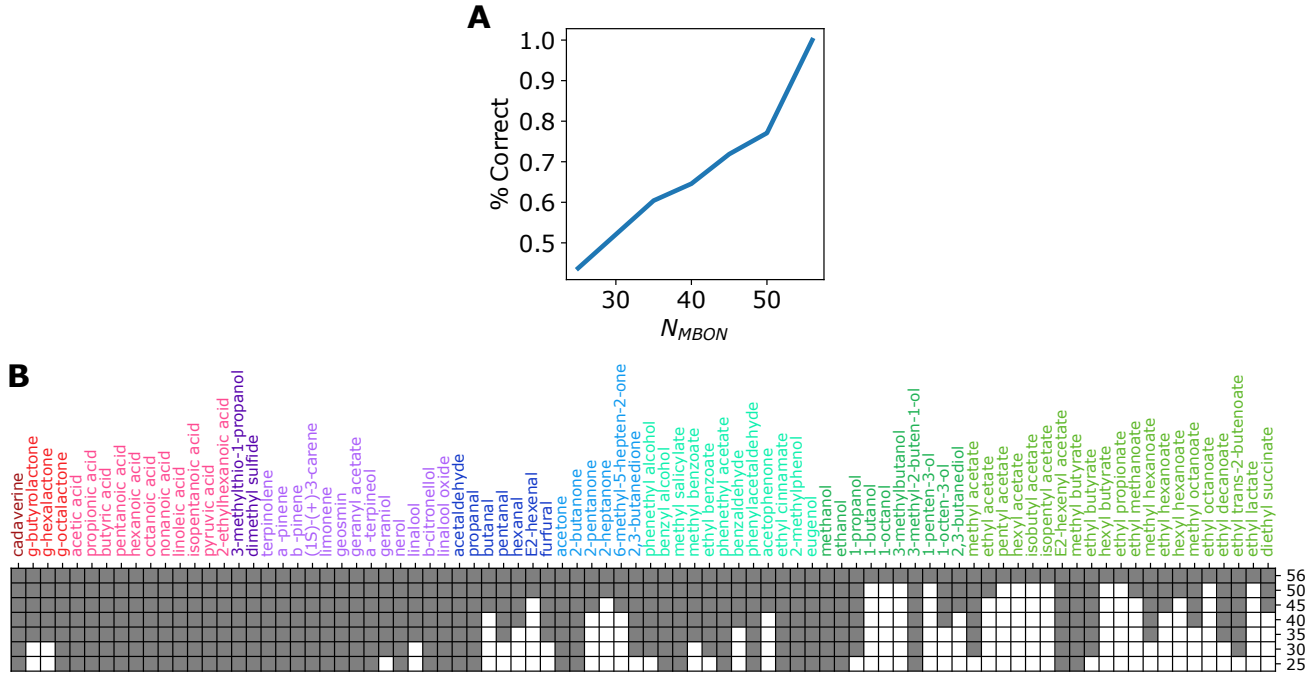


Figure S12: **Testing the odor identifiability using a random subset of MBONs of the size N_{MBON}.** (A) The relationship between N_{MBON} and the success rate of odor identification. For a given odorant and a given size N_{MBON}, MBONs are sampled ten times, where the odor identification is considered successful if at least one set of the sampled MBONs can accurately identify the given olfactory stimulus. We employ the 86 odorants where the identification was successful in the original setup. (B) The results of the identification test on 86 odorants from a subset of MBONs with the size N_{MBON} labeled on the y-axis. Gray denotes success and white denotes failure.