J Comput Neurosci (2011) 30:143–161
DOI 10.1007/s10827-010-0265-0

System identification of *Drosophila* olfactory sensory neurons

Anmo J. Kim · Aurel A. Lazar · Yevgeniy B. Slutskiy

Received: 6 January 2010 / Revised: 18 July 2010 / Accepted: 26 July 2010 / Published online: 21 August 2010

Abstract The lack of a deeper understanding of how olfactory sensory neurons (OSNs) encode odors has hindered the progress in understanding the olfactory signal processing in higher brain centers. Here we employ methods of system identification to investigate the encoding of time-varying odor stimuli and their representation for further processing in the spike domain by *Drosophila* OSNs. In order to apply system identification techniques, we built a novel low-turbulence odor delivery system that allowed us to deliver airborne stimuli in a precise and reproducible fashion. The system provides a 1% tolerance in stimulus reproducibility and an exact control of odor concentration and concentration gradient on a millisecond time scale. Using this novel setup, we recorded and analyzed the *in-vivo* response of OSNs to a wide range of time-varying odor waveforms. We report for the first time that across trials the response of OR59b OSNs is very precise and reproducible. Further, we empirically show that the response of an OSN depends not only on the concentration, but also on the rate of change of the odor concentration. Moreover, we demonstrate that a two-dimensional (2D) Encoding Manifold in a concentration-concentration gradient space provides a quantitative description of the neuron's response. We then use the white noise system identification methodology to construct one-dimensional (1D) and two-dimensional (2D) Linear-Nonlinear-Poisson (LNP) cascade models of the sensory neuron for a fixed mean odor concentration and fixed contrast. We show that in terms of predicting the intensity rate of the spike train, the 2D LNP model performs on par with the 1D LNP model, with a root mean-square error (RMSE) increase of about 5 to 10%. Surprisingly, we find that for a fixed contrast of the white noise odor waveforms, the nonlinear block of each of the two models changes with the mean input concentration. The shape of the nonlinearities of both the 1D and the 2D LNP model appears to be, for a fixed mean of the odor waveform, independent of the stimulus contrast. This suggests that white noise system identification of Or59b OSNs only depends on the first moment of the odor concentration. Finally, by comparing the 2D Encoding Manifold and the 2D LNP model, we demonstrate that the OSN identification results depend on the particular type of the employed test odor waveforms. This suggests an adaptive neural encoding model for Or59b OSNs that changes its nonlinearity in response to the odor concentration waveforms.

Keywords System identification · Olfactory sensory neurons · White noise analysis · I/O modeling

Action Editor: C. Linster

Electronic supplementary material The online version of this article (doi:10.1007/s10827-010-0265-0) contains supplementary material, which is available to authorized users.

The authors' names are alphabetically ordered.

A. J. Kim · A. A. Lazar (✉) · Y. B. Slutskiy
Department of Electrical Engineering,
Columbia University, New York, NY, USA
e-mail: aurel@ee.columbia.edu

Alphabetical list of abbreviations

1D	One-dimensional
2D	Two-dimensional

BARS	Bayesian Adaptive Regression Splines
DPG	Dipropylene glycol
I/O	Input/output
LNP	Linear-nonlinear-Poisson
MID	Maximally informative dimensions
OSN	Olfactory sensory neuron
PID	Photoionization detector
PSTH	Peristimulus time histogram
RCO	Reverse correlation
RMSE	Root-mean-square error
STA	Spike-triggered average
STC	Spike-triggered covariance

1 Introduction

A functional characterization of an unknown system typically begins by making observations about the response of that system to input signals. The knowledge obtained from such observations can then be used to derive a quantitative model of the system in a process called system identification (Wu et al. 2006; Marmarelis 2004). More precisely, the goal of system identification is to use a given input/output data to derive a function that maps an arbitrary system input into an appropriate output.

In neurobiology, system identification has been applied to a variety of systems, ranging from the ion channels of the giant squid axon to the visual pattern recognition circuits in the Macaque monkey (Hodgkin and Huxley 1952; Rust et al. 2006). Depending on the level of abstraction, the identified neural models vary from detailed mechanistic models (Lindemann 2001; Rospars et al. 2003; Dougherty et al. 2005; Reidl et al. 2006; Gu et al. 2009; Kaissling 2009; Halnes et al. 2009) to purely phenomenological models (Marmarelis and Naka 1972; Marmarelis 2004; Pillow and Simoncelli 2006).

System identification techniques usually follow three steps. First, a mathematical model of a given system is hypothesized. This model can be either generic (Marmarelis 2004) or restricted to a certain architecture (Rust et al. 2006). Second, parameters of the hypothetical model are estimated using the input/output data of the system. Typically, the parameter estimation is posed as an optimization problem in which a certain measure of performance is either minimized or maximized. Examples of performance measures include the mean-square error, the likelihood function, and the mutual information. Finally, in a cross-validation step, the performance of the proposed model is evaluated using previously unseen data.

The estimation and cross-validation steps of the system identification call for a careful design of input stimuli as well as their precise control and measurement. This has been possible in many sensory systems, including vision and audition, where system identification of a variety of neural circuits has been carried out under the assumption that they are in a steady state (Marmarelis and Naka 1972; Hunter and Korenberg 1986; Steveninck and Bialek 1988; Brenner et al. 2000; Slee et al. 2005). In olfaction, however, little progress has been made primarily because of the difficulties associated with the control, measurement and reproducible delivery of odor stimuli.

In this paper, we employ a novel odor delivery and measurement system to record the extracellular activity of individual *Drosophila* OSNs in response to airborne odor stimuli. We use system identification techniques to investigate the nature of the encoding and spike-domain representation of odorants by OSNs. It is our firm belief that the lack of such fundamental research has hindered the progress in understanding the olfactory signal processing in higher brain centers. To our best knowledge, this study represents the first attempt to apply system identification methodology to *in-vivo* OSN recordings in *Drosophila* Melanogaster.

The rest of the paper is organized as follows. In Section 2 we describe our experimental methods, including the odor delivery system. In Section 3 we present our initial recordings and an empirical I/O modeling of OSNs. In Section 4 we discuss the system identification of OSNs using white noise input stimuli. There we describe and compare the 1D and 2D Linear-Nonlinear-Poisson models of *Drosophila* OSNs. Finally, the paper concludes with a discussion in Section 5.

2 Experimental methods

2.1 Odor delivery system

We built a novel low-turbulence odor delivery system that allowed us to deliver airborne odorants in a precise and reproducible fashion. The system provides an exact control of odor concentration and concentration gradient on a millisecond time scale.

2.1.1 System design

The block diagram of our odor delivery system is shown in Fig. 1. Compressed medical air is first humidified using distilled water and then split into two

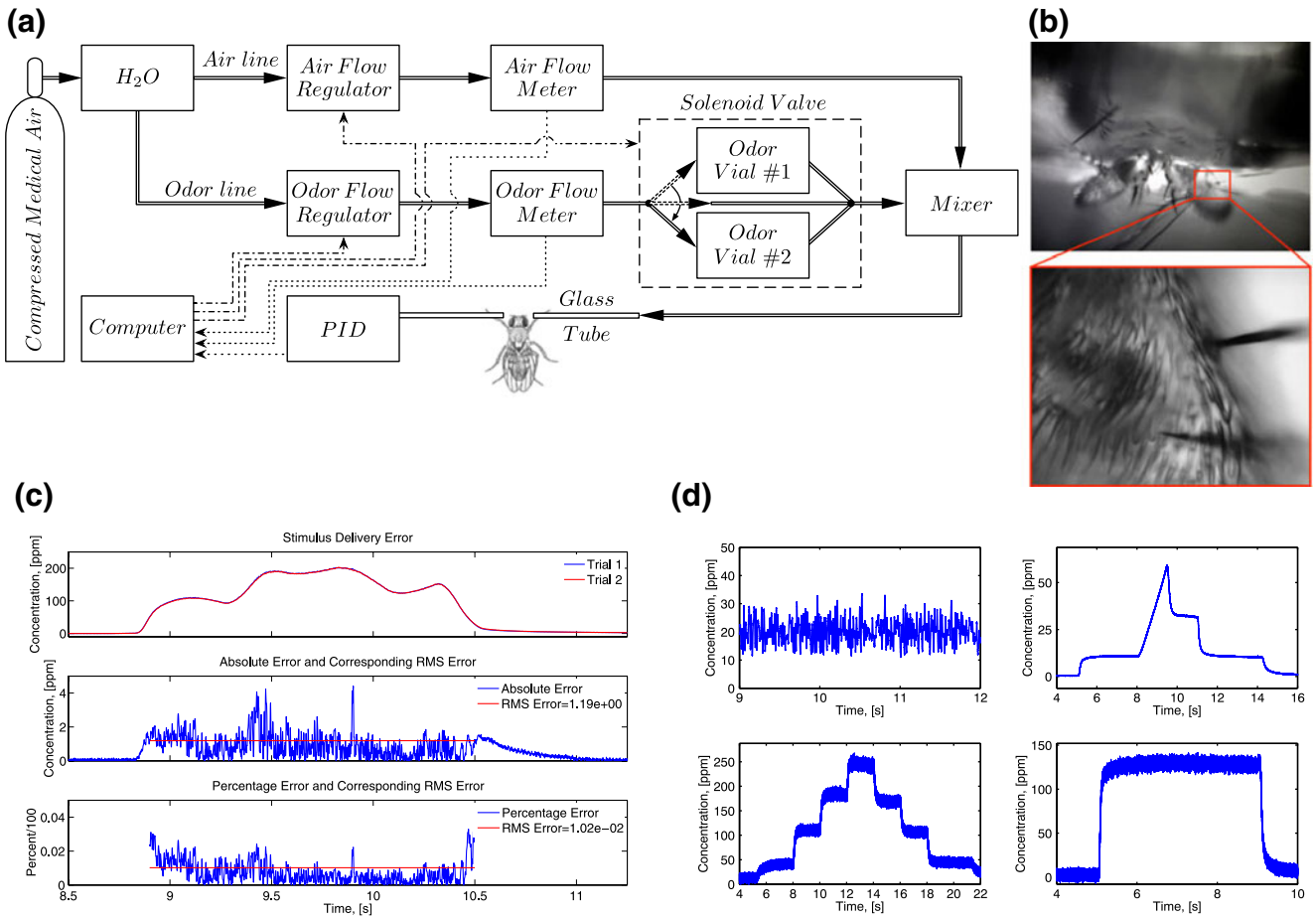

Fig. 1 The experimental setup. (**a**) The block diagram of the odor delivery system. (**b**) *Top*: the fly head and antennae as seen under the microscope; *bottom*: two electrodes are placed into two distinct sensilla for simultaneous *in-vivo* extracellular recordings. (**c**) The tolerance of the stimulus reproducibility is 1%. (**d**) The setup allows to deliver a variety of odor concentration waveforms

airstreams: one for an odor line and another for an air line. The airstream in the air line passes through a flow regulator, a flow meter and enters a laminar mixer (laminarity of the flow was evaluated indirectly through odor concentration measurements, data not shown). Similarly, the airstream in the odor line passes through a flow regulator and a flow meter. However, before entering the laminar mixer, the airstream in the odor line is redirected to a three-way solenoid valve. Depending on the setting of the valve, the odor airstream goes either directly to the laminar mixer or passes through one of the odor vials connected to the valve. If going through one of the vials, the airstream picks up odor molecules in a given vial and enters the laminar mixer, where it is combined with the airstream from the air line. The combined air/odor-stream is then delivered to the antennae of a fruit fly through a glass tube. Directly opposite the glass tube, a photoionization detector (PID) takes in the air surrounding the fly antennae and measures the odorant concentration. The PID intake rate is 1 L/min and the combined air/odor-stream is delivered at a rate of 700–800 mL/min.

A computer is used to control the flow regulators in both the air line and the odor line as well as the opening and closing of the 3-way solenoid valve. Measurements of the flow in both lines were used in a feedback mechanism to adjust the flow regulator values. Further, the output of both flow meters and that of the PID was recorded and analyzed offline to obtain the odor concentration. Because the sensitivity of the PID drops gradually in time, an additional odorant (Ethyl Acetate or Hexane) is delivered periodically at a predetermined concentration to recompute the transfer function of the device. This transfer function is used to convert the voltage output of the PID (Volts, [V]) into the odorant concentration (parts per million, [ppm]).

2.1.2 System performance

Using a photoionization detector, we were able to measure the concentration of the delivered odorant in real time and check the reproducibility of the setup. We found that odor waveforms reaching the antennae of fruit flies can be reproduced to within a tolerance of 1% (Fig. 1(c)). The system also allowed us to deliver a variety of time-varying odor concentration waveforms, some of which are shown in Fig. 1(d).

2.2 Stimulus design

2.2.1 Odorant preparation

Odorants used in this study were obtained from Sigma-Aldrich in liquid form at high purity. A half an hour before the experiment, an odorant was poured into a 30 mL glass vial, diluted with dipropylene glycol (DPG) and sealed with a rubber stopper and a plastic screw-on cap. Miscibility of all odors with DPG was thoroughly checked before the experiments.

2.2.2 White noise odor stimuli

White noise odor stimuli were produced by modulating the flow rate in both the air line and the odor line (Fig. 1(a)) around fixed-mean flow rates. Both rates were modulated by sending a sequence of independent identically distributed control values to the corresponding flow regulator. Each control value was picked from a Gaussian distribution and updated every millisecond. Such a frequent update guaranteed that control actions were taken within the minimal response time of the flow regulators. While airflow-dependent, the minimal response time of each flow regulator was below 10 ms.

Using the above protocol, we generated pseudo-white Gaussian noise odor waveforms that exhibit a nearly flat power spectral density of up to 30 Hz (see Figs. S1–S3 in the Supplemental material). Such waveforms are sufficiently white to perform system identification since the bandwidth of insect olfactory systems typically does not exceed 25 Hz: 20 Hz in the *American cockroach*, 20 Hz in three moth species and 2 Hz in the *American lobster* (Lemon and Getz 1997; Bau et al. 2002; Gomez et al. 1999). In our own studies (data not shown) we estimated the bandwidth of *Drosophila* OSNs to be roughly 25 Hz.

One example of a Gaussian noise odor waveform (sample path) is the red trace shown in Fig. 4(a). Note the time scale of the odor waveform. The noise sequence is 8 s long and is applied 1 s after the odor onset so as to let the neuron reach its steady-state response. The total duration of the waveform is 9 s.

2.2.3 Triangle odor waveforms

Triangle odor waveforms were produced by fixing the flow in the air line and first slowly increasing and then slowly decreasing in a controlled fashion the flow in the odor line (Fig. 1(a)). We designed a set of nine triangle waveforms so that in combination with different ligand dilutions we could explore a large range of odor concentration and concentration gradient (rate of change) values. An example of such a set of odor waveforms is shown in Fig. 3(a).

Note the timescale of the odor waveforms in Fig. 3(a). Each triangle waveform was designed to be roughly 2 s long to allow enough time for the transient response of the neuron, while also limiting the effect of adaptation.

2.3 *Drosophila* stocks

Drosophila stocks were maintained at a room temperature on a 12-h light/12-h dark schedule and kept in standard plastic vials containing a cornmeal-agar medium. All of the experiments were performed on female wild-type (*Canton-S*) flies three-to-five days post-eclosion.

2.4 Electrophysiology

Electrophysiological methods in this study are similar to those previously described in Clyne et al. (1997).

2.4.1 Fly preparation

A female fly was prepared for *in-vivo* recordings a half an hour prior to each experiment. The fly was taken out of the vial and placed into a plastic micropipette tip with the head of the fly facing the narrow end of the tip. The tip was cut just a few millimeters behind the fly body and a small amount of industrial plasticine was placed to prevent the fly from escaping. Another cut was made 1 mm before the anterior side of the head. Forceps were used to gently push the fly in order to fully expose the antennae and only partially the eyes at the front end of the micropipette tip. Special care was taken so as not to damage the animal. Next, the micropipette tip with the fly was placed on a glass slide with a stack of glass coverslips glued to it. The tip was attached to

the slide using industrial plasticine with the antennae of the fly laying flat on the coverslip stack. The slide was then placed onto a microscope (Eclipse E600FN, Nikon) that was mounted on a vibration isolation table (63–500 Series, Technical Manufacturing). Finally, a glass electrode attached to a micro-manipulator (MX10R, Siskiyou) was used to rotate one of the antennae into a desired position and to immobilize it.

2.4.2 Neural recordings

Three tungsten electrodes were electrolytically sharpened using a 5% dilution of potassium hydroxide and mounted on separate motorized micro-manipulators (MP-285, Sutter Instrument). The electrodes were connected to a preamplifier (MultiClamp 700B, Axon Instruments/Molecular Devices) and their output was band-pass filtered with cutoff frequencies $f_{\text{low}} = 30$ Hz and $f_{\text{high}} = 2$ kHz. The output of the amplifier was fed to a data acquisition system (Digidata 1322A, Axon Instruments/ Molecular Devices) and then to a computer so that the activity of neurons could be monitored in real time using the pClamp software (Axon Instruments/Molecular Devices). The amplifier output was also stored for further analysis offline.

The tip of one of the electrodes was inserted into the compound eye of the fly and used for signal ground. The other two electrodes were used to simultaneously record the extracellular activity of neurons from two different sensilla. The tip of each recording electrode was inserted at the base of a sensillum in order to make contact with the conducting sensillum lymph. The spontaneous activity of the neurons was checked against that reported in de Bruyne et al. (2001) to ensure that the neurons were not damaged. If a neuron was observed to suddenly increase its spiking activity in the absence of odorants, it was presumed to be damaged and recordings from that neuron were discontinued.

All recordings in this study were taken from Or59b olfactory sensory neurons that are located in the ab2 large basiconic sensilla. Action potentials were sorted and analyzed using custom software in MATLAB.

3 Initial recordings and empirical modeling

We recorded the neural activity of ab2A neurons expressing the Or59b receptor in response to a variety of time-varying odor concentration waveforms. These particular neurons were chosen because they respond to acetone, an odor that is well ionized and consequently easily detected by the PID.

3.1 OSN response to a staircase waveform

One of the time-varying waveforms used in our study was the staircase waveform, shown in Fig. 2(a). In this waveform the odor concentration is incremented in steps of roughly 40 ppm until the maximum concentration of 110 ppm is reached. Then the concentration is decreased in the same step-like fashion, with each step being 2 s long. In Fig. 2(b) we plot the spike raster of the neuron's response, while in Fig. 2(c) we show the corresponding PSTH. Note that the concentration of the delivered odor waveforms in the repeated experiments is not identical. To highlight this departure from similar experiments in other sensory systems (e.g., vision and audition), where the delivery of identical stimuli is straightforward, we use a different color to mark each trial in Fig. 2(b).

For greater clarity, the same raster plot is also shown in black and white in Fig. S5(B) of the Supplemental material.

Note that whenever there is a sudden positive increment in the concentration, the instantaneous firing rate of the neuron increases dramatically. For example, when the odor concentration is increased by 40 ppm at time $t = 10$ s, the spike rate of the neuron jumps from 40 Hz to 70 Hz. At the same time, whenever there is a negative change in the concentration, the instantaneous firing rate of the neuron goes down. For example, this can be clearly seen at time $t = 14$ s when the concentration falls by roughly 40 ppm and the spike rate of the neuron drops from about 55 Hz to 35 Hz.

We also note that the firing rate of the neuron depends on the mean of the odor concentration waveform. This can be clearly seen after the transient response of the neuron dies away. For example, for time $t \in [9\,\text{s}, 10\,\text{s}]$ the odor concentration is constant and is equal to roughly 40 ppm. Then the concentration is increased and eventually reaches a constant concentration of 80 ppm for time $t \in [11\,\text{s}, 12\,\text{s}]$. At the same time, the firing rate of the neuron goes up from 40 Hz to 50 Hz.

Concluding, in response to acetone, the Or59b OSNs detect and encode both the mean and the temporal changes of the odor concentration.

3.2 OSN responses to triangle odor waveforms

To get a better understanding of how the odor concentration and temporal changes in the odor concentration affect the OSN response, we designed a set of nine triangle odor waveforms shown in Fig. 3(a). In combination with different ligand dilutions, this set

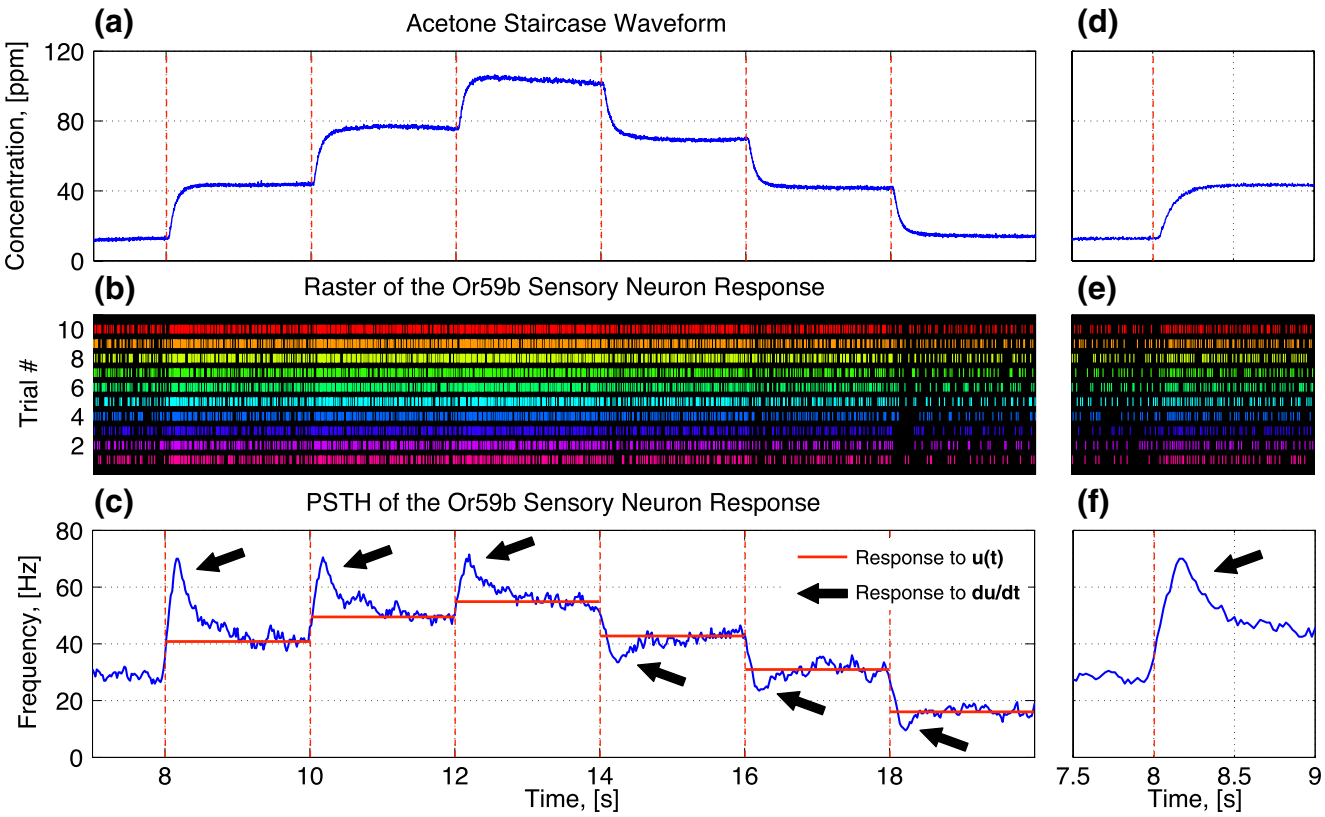

Fig. 2 The OSN response to the staircase waveform. (**a**) The staircase acetone odor waveform is plotted against time. The *dashed vertical lines* (*red*) indicate the times at which the odor concentration is either increased or decreased in a step-like fashion. The length of each step is 2 s. (**b**) The raster of the OSN response to 10 consecutive presentations of the staircase odor waveform. A different color is used for each trial # to highlight that in repeated trials the delivered odor waveforms differ from each other by up to 1%. (**c**) The PSTH of the OSN response to the staircase waveform was computed using a 100 ms bin size with a 25 ms sampling interval. *Red horizontal lines* denote the OSN response to the odor concentration and black arrows point out the neural response to the rate of change of the odor concentration. (**d**)–(**f**) A one-second-long window from (**a**)–(**c**)

of waveforms allowed us to explore a large range of odor concentration and concentration gradient (rate of change) values. Thus we could parametrize the response of the neuron in terms of the odor concentration and its rate of change and obtain a quantitative description of their combined effects.

In Fig. 3(b) we demonstrate the OSN response to three different triangle waveforms in Fig. 3(a). We use the same color code to show the PSTH of the neural response to triangles #1 (red), #5 (green), and #9 (blue). The PSTH was computed with a 100 ms bin size and a 25 ms sampling interval. Note that the three PSTHs are very distinct and depend on the temporal properties of the waveforms. In all cases however, the response clearly varies with the odorant concentration. Moreover, as already expected, the neural response also depends on the rate of change of the odor concentration. The highest and lowest maximal spike rates are generated in response to triangles #1 and #9, respectively. This reflects their higher and lower onset gradients, when compared to the triangle waveform #5.

3.3 The 2D Encoding Manifold

In Fig. 3(c) we plot the triangle odor waveforms in the concentration-concentration gradient plane for multiple odorant dilutions. Note that the concentration values vary from 0 ppm to 250 ppm while the concentration gradient values vary from roughly −1500 ppm/s to +2250 ppm/s. We use the same colors as in Fig. 3(a) and (b) to highlight the resulting trajectories of the triangle waveforms #1, #5, and #9. We note that at time $t = 8.5$ s the concentration of the three waveforms in Fig. 3(a) is $u = 0$ and the rate of the concentration change is $\dot{u} = 0$. This corresponds to the point $(0, 0)$ in Fig. 3(c). With time, the odor concentration increases in Fig. 3(a) up to a peak of roughly 210 ppm. In Fig. 3(c) this corresponds to the movement along the highlighted

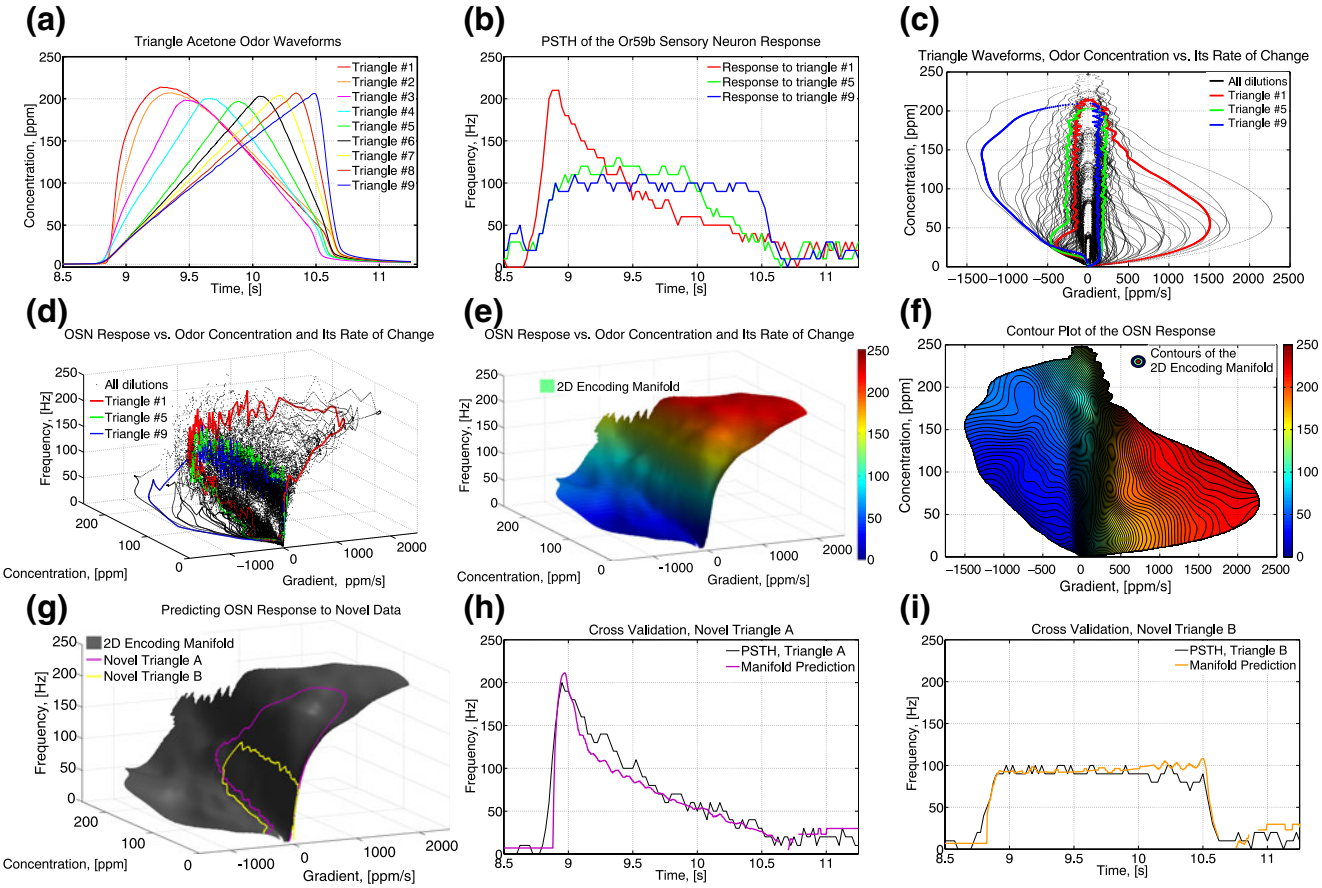

Fig. 3 The triangle odor waveforms and the 2D Encoding Manifold. (**a**) Nine triangle acetone odor waveforms for a single odorant dilution are plotted against time. (**b**) The PSTH of the Or59b OSN response to triangle waveforms #1 (*red*), #5 (*green*) and #9 (*blue*) was computed using a 100 ms bin size and a 25 ms sampling interval. (**c**) The trajectories of triangle waveforms for all dilutions are plotted in the concentration-concentration gradient (rate of change) plane. The trajectories of triangles #1, #5 and #9 are shown using the same colors as in (**a**), (**b**). (**d**) The neural response to all triangle waveforms is plotted as a function of the concentration and its rate of change. The trajectories of the response to triangles #1, #5, and #9 are shown using the same colors as in (**a**)–(**c**). (**e**) The 2D Encoding Manifold is generated by applying a 2D ridge estimator to the data in (**d**). (**f**) The contour plot of the manifold in (**e**). (**g**) Cross-validation: the 2D Encoding Manifold is used to read out the OSN response to two novel triangle odor waveforms. (**h**), (**i**) The read-out response (color) is plotted against time and compared with the PSTH (*black*) of the OSN response

trajectories in the counter-clockwise direction through the right half-plane (positive gradient). Further counter-clockwise movement through the left-half plane (negative gradient) corresponds to the falling side of the triangle odor waveforms in Fig. 3(a). At time $t = 11$ s the concentration of the three waveforms is again $u = 0$ in Fig. 3(a) and the rate of the concentration change is again $\dot{u} = 0$. This corresponds to all trajectories coming back to the point $(0, 0)$ in Fig. 3(c).

Next, we parametrize the OSN response to triangle waveforms at all dilutions using the odor concentration and its rate of change. At each time step, the PSTH of the neural response ([Hz]) is associated with the corresponding values of the odor concentration ([ppm]) and its rate of change ([ppm/s]). We thus obtain a three-dimensional description of the input/output data (Fig. 3(d)). In Fig. 3(e) we show the surface produced by applying a 2D ridge estimator (see Sec. 2.3 of the Supplemental material) to the data in Fig. 3(d). And in Fig. 3(f) we show the contour plot of the same surface. We call this surface the 2D Encoding Manifold since it provides a quantitative description of the odor encoding performed by an OSN.

Examining Fig. 3(e) and (f), we note that the 2D Encoding Manifold is highly nonlinear and that the Or59b OSN clearly encodes the information about both the odor concentration and its rate of change. The

neuron responds very strongly to even the smallest positive values of the gradient and encodes only positive concentration gradients at low odor concentrations. At high concentrations the OSN mostly encodes the odor concentration.

To demonstrate that the above manifold quantitatively captures the encoding properties of an OSN, we use cross-validation. We calculate the values of the odor concentration and its rate of change for two novel triangle waveforms. At every time step, we use these values to read out the value of the neural response from the manifold. This process is visualized in Fig. 3(g). The response to the novel triangles 'A' (magenta) and 'B' (orange) corresponds to two different trajectories on the 2D Encoding Manifold (black). We plot the readout response values as a function of time in Fig. 3(h) and (i). Note that the predicted response (color) in both figures closely follows the PSTH (black) of the neuron.

Thus, at the very least for this class of triangle odor waveforms, the OSNs encode information about the odor concentration and its rate of change. This is remarkable as it demonstrates a sophisticated processing and representation of olfactory information at the very first layer of the olfactory system.

4 Identification of OSNs using white noise stimuli

In contrast to the empirical modeling presented above, a principled system identification approach provides a clear set of guidelines for combining the input/output data with other knowledge about the system to estimate a function that maps the input space into the output space of the system. In order to obtain an estimate of this function, it is necessary to choose a computational model that can provide a good description of the system.

Quantitative neurophysiological studies have produced several classes of computational models describing the functional relationship between sensory inputs and neural responses. The model classes differ in their assumptions about the properties of a neuron or a neural circuit and impose constraints on the choice of the optimal mapping function (Wu et al. 2006).

4.1 Overview of the LNP cascade model

One class of computational models that became popular in sensory neurophysiology is the Linear-Nonlinear-Poisson class of models. Originally proposed by Wiener (1958), this class has been successfully used to model numerous neural circuits in vision, audition and vestibular systems (Marmarelis and Naka 1972; Victor and Shapley 1980; Aertsen and Johannesma 1981; Hunter and Korenberg 1986).

In its simplest form, the LNP model consists of (i) a static linear block, or a filter, that performs linear processing on an input stimulus and describes how the input stimulus is converted into the intracellular voltage; (ii) a static nonlinear block that maps the output of the filter to a spike intensity rate, taking into account such nonlinearities as rectification and saturation; and (iii) a Poisson block that generates a train of spikes as an inhomogeneous Poisson process.

Even though the LNP cascade model is a phenomenological model and does not take into account many biophysical details (such as the spike generation), it often provides a compact and reasonably accurate description of average neural responses (e.g., the PSTH) in many early sensory areas (Pillow 2007).

Formally, the static linear block of the LNP model consists of a set of n fixed linear filters $\{h_i\}_{i=1}^n$ that are fully described by their kernels $h_i(t)$, $t \in \Re$, $i = 1, \ldots, n$. Let $u(t)$, $t \in \Re$, be a stimulus at the input to a system. Then the input to the linear block of the LNP model is a zero-mean stimulus $(u - \bar{u})(t)$, $t \in \Re$, $\bar{u} = \lim_{T\to\infty} \frac{1}{T} \int_0^T u(s)ds$. At any time t, the output of the linear block is an n-dimensional vector $\mathbf{v} = [v_1(t), \ldots, v_n(t)]$, where

$$v_i(t) = \big(h_i * (u - \bar{u})\big)(t) = \int_{\Re} h_i(s)\big[u(t-s) - \bar{u}\big]ds,$$

and $*$ denotes the convolution.

The output $\mathbf{v} = [v_1(t), \ldots, v_n(t)]$ of the linear block feeds into an n-dimensional nonlinearity to produce the random intensity rate $\lambda(t)$, $t \in \Re$, of a conditional (doubly stochastic) Poisson process:

$$\lambda(t) = f(\mathbf{v}) = f\big(v_1(t), \ldots, v_n(t)\big),$$

where $f : \Re^n \to \Re$ models the nonlinear block.

If there is only one filter in the linear block of the LNP model, then the nonlinear block is 1-dimensional and the model is called the 1D LNP model. Similarly, if the linear block consists of two filters, then the model is called the 2D LNP model. Typically, only the 1D and 2D LNP models are used to describe the response of a neural circuit since the amount of data required to estimate the parameters of higher-order models scales exponentially with the order of the model (Victor 2005).

4.2 Estimation of the LNP model parameters

The system identification problem is effectively a regression problem: given a limited input/output data

of the system, identify a function that maps an arbitrary input to an appropriate output. To solve this regression problem, many different parameter estimation methods with different constraints and optimality criteria have been developed. For the linear block we will only discuss the reverse correlation (RCO), spike-triggered average (STA), maximally informative dimensions (MID) and the spike-triggered covariance (STC) methods. For the nonlinear block, ridge and polynomial regression methods are employed.

4.2.1 Estimation of the 1D linear block

The reverse correlation method for the estimation of the 1D linear block is based on a result of Bussgang (1952) and was first employed by Hunter and Korenberg (1986). In this method, the impulse response of the linear filter h_{RCO} is obtained using the reverse correlation (RCO) between the zero-mean stimulus $u(t) - \bar{u},\ t \in \Re$, and the estimate $\hat{\lambda}$ of the random intensity rate λ of the spike train:

$$h_{\text{RCO}}(t) = \int_{\Re} (u(s) - \bar{u})\hat{\lambda}(s-t)ds. \tag{1}$$

Typically, the PSTH of the neural response is used as the estimate $\hat{\lambda}$ with

$$\hat{\lambda}_{[a,b]} = \frac{1}{b-a}\frac{1}{K}\sum_{j=1}^{K} N^j_{[a,b]}, \tag{2}$$

where $N^j_{[a,b]}$ is the spike count on the time interval $[a, b]$ for the j^{th} presentation of the stimulus.

An alternative method to estimate the linear block of the 1D LNP model is called the spike-triggered average. As the name suggests, the STA filter $h_{\text{STA}} = h_{\text{STA}}(t), t \in \Re$, is computed by averaging fixed-length segments of the mean-zero input stimulus directly preceding every spike in the spike train:

$$h_{\text{STA}}(t) = \lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m} u(t_k - t) - \bar{u}, \tag{3}$$

where $t \in [0, S]$, S is the length of the impulse response of the filter, and (t_k), $k = 1, \ldots, m$, represents the sequence of spike times. When applying the STA method, white Gaussian noise is typically used at the input.

Yet another method to estimate the linear block, called maximally informative dimensions, was described in Sharpee et al. (2004) and Paninski (2003). The MID method estimates the linear kernel by maximizing a measure between the *prior* and the spike-triggered distribution of the filter output. This method does not impose any conditions on the distribution of an input stimulus; non-Gaussian and non-white inputs can be used. Typically, the linear filter $h_{\text{MID}} = h_{\text{MID}}(t), t \in \Re$, is estimated by maximizing the Kullback–Leibler divergence, i.e.,

$$h_{\text{MID}}(t) = \arg\max_{h} D, \tag{4}$$

where D is a distance measure between the probability distribution of the filter output $v(t) = (h * (u - \bar{u}))(t)$ and the probability distribution of the filter output $v(t)$ conditioned on the time of a spike (Pillow and Simoncelli 2006).

4.2.2 Estimation of the 2D linear block

The MID method can be extended to maximize the Kullback–Leibler divergence with respect to two different directions and thus estimate kernels h_1 and h_2 of the 2D LNP model (Sharpee et al. 2004).

Alternatively, the two filters of the 2D LNP cascade model can be estimated from the information provided by the spike-triggered covariance matrix (Steveninck and Bialek 1988; Brenner et al. 2000; Slee et al. 2005; Fairhall et al. 2006; Schwartz et al. 2006; Geffen et al. 2009). In this approach, the prior stimulus and the spike-triggered stimulus are compared by evaluating their covariances. The two covariance matrices are defined as follows:

$$C_P(t,s) = \mathbf{E}[u(\tau - t)u(\tau - s)] - \mathbf{E}[u(\tau - t)]\mathbf{E}[u(\tau - s)]$$

$$C_{STC}(t,s) = \mathbf{E}[u(\tau - t)u(\tau - s)|\tau = t_k] - \\ -\mathbf{E}[u(\tau - t)|\tau = t_k]\mathbf{E}[u(\tau - s)|\tau = t_k],$$

where $\mathbf{E}[\cdot]$ represents the mathematical expectation and t_k is an arbitrary spike time.

In practice, the input signal is discretely sampled, and thus both C_{STC} and C_P are finite-dimensional matrices. The kernels $h_1(t)$ and $h_2(t)$ of the two linear filters of the 2D LNP cascade model are the eigenvectors corresponding to the two most significant eigenvalues of the matrix

$$C = C_{\text{STC}} - C_P, \tag{5}$$

where C can be analyzed using eigenvalue decomposition (Steveninck and Bialek 1988). The magnitude of the eigenvalues represents the change in variance from the prior stimulus distribution to the STA output distribution along the corresponding eigenvectors.

4.2.3 Estimation of the nonlinearity

For both the 1D and the 2D LNP model, the nonlinear block is estimated by simply matching the output of the linear block $\mathbf{v} = [v_1(t), \ldots, v_n(t)]$, $n = 1$ or $n = 2$, to the

estimate $\hat{\lambda}$ of the random intensity rate λ. Typically, this is done by minimizing the residual

$$M_{[a,b]} = N_{[a,b]} - \int_a^b \lambda(s)ds = N_{[a,b]} - \int_a^b f(v(s))ds, \tag{6}$$

where $N_{[a,b]} = \frac{1}{K}\sum_{j=1}^{K} N^j_{[a,b]}$ is the spike count estimate on the time interval $[a, b]$ obtained from K trials and the residual $M_{[a,b]}$ is attributed to the (white) noise in the neural response. Ridge and polynomial regression methods are used for obtaining an estimate of f in Eq. (6) (see Section 2.2 in the Supplemental material for details).

4.3 LNP Models of the Or59b OSNs

Here we employ LNP cascade models to identify the Or59b OSN in response to white noise waveforms with fixed mean and variance values. Both frozen and white noise experiments are carried out to estimate a 1D linear kernel and the corresponding nonlinearity block. A 2D LNP cascade model is also estimated from the same set of data. The estimated systems are cross-validated using an independent white noise sample path with the same statistical parameters as the training (or test) waveforms.

4.3.1 OSN response to frozen noise odor waveforms

In order to estimate parameters of any phenomenological neuron model, precise measurements of the input stimulus and the output spike train must be available to an observer. In olfaction, this has not been possible until now primarily because of the difficulties associated with the reproducible delivery and accurate measurement of the odorant concentration. We note that the experimental reproducibility is absolutely essential if one wishes to use the PSTH and/or Bayesian Adaptive Regression Splines (BARS) to estimate the instantaneous firing rate $\lambda(t), t \in \Re$, of a neuron in response to a given odor waveform (Kass et al. 2003, 2005).

In Fig. 4 we demonstrate that our experimental setup allows us to take precise measurements of both the input and the output. A reproducible delivery of a frozen noise odor waveform results in a precise and structured response of the OSN. Figure 4(a) shows the time course of 60 repeated frozen noise acetone odor waveforms that were delivered to the antennae of a fly. All waveforms were generated using the same control sequence and are practically identical. Figure 4(d) shows a 1 s window from Fig. 4(a) and demonstrates the high degree of reproducibility of odor waveforms in all 60 repeated trials.

In Fig. 4(b) we plot the raster of the Or59b OSN response to 60 waveforms shown in Fig. 4(a). Every row in the plot corresponds to a single trial and each vertical line segment denotes an action potential. We use a different color for each trial to highlight that the delivered odor waveforms in repeated trials differ from each other by up to 1%.

We note that such a precise and reproducible odor delivery system allows one to observe the remarkable spiking precision of the OSN. For instance, the start and the end of the neural response in Fig. 4(b) are clearly correlated with the onset and the offset of the odor concentration waveform. Furthermore, the clear vertical gaps in the raster are correlated with the noisy fluctuations of the odor concentration. This data suggests that the response of an OSN is precise and highly reproducible. To our knowledge, this has not been observed before primarily because of the difficulties associated with the odor delivery and measurement.

Finally, in Fig. 4(c) we show the instantaneous firing rate of the neuron. First, we plot the instantaneous firing rate using a PSTH (blue) with a bin size of 20 ms and a sampling interval of 1 ms (hence the "20/1" notation). Second, we apply BARS (red) to a PSTH with non-overlapping 10 ms bins (hence the "10/10" notation). This allows us to obtain an additional estimate of the instantaneous firing rate.

4.3.2 The 1D LNP model of the Or59b OSN

Figure 5(a) depicts the LNP cascade model of an Or59b OSN in response to acetone waveforms. The model consists of a linear filter (L Block) followed by a static nonlinear block (N Block). A zero-mean input $u(t) - \bar{u}$ is provided at the input to the model and the intensity rate $\lambda(t)$ of the spike train is read out at the output.

The linear block in Fig. 5(a) was estimated using all three of the aforementioned techniques: RCO, STA and MID. All three filter shapes are akin to each other except that the MID filter exhibits a sharper peak as well as noisy fluctuations. The nonlinearity in Fig. 5(a) was computed using the first order polynomial fit (green dotted line) as well as a ridge estimator (red line) discussed in Sec. 2.2 of the Supplemental material.

The estimated 1D LNP cascade model is cross-validated by measuring the prediction error for an arbitrary white noise input sample path drawn from the Gaussian distribution with mean $\mu = 66$ ppm and variance $\sigma^2 = 25^2$. The prediction error is scored by computing the RMSE values between the measured

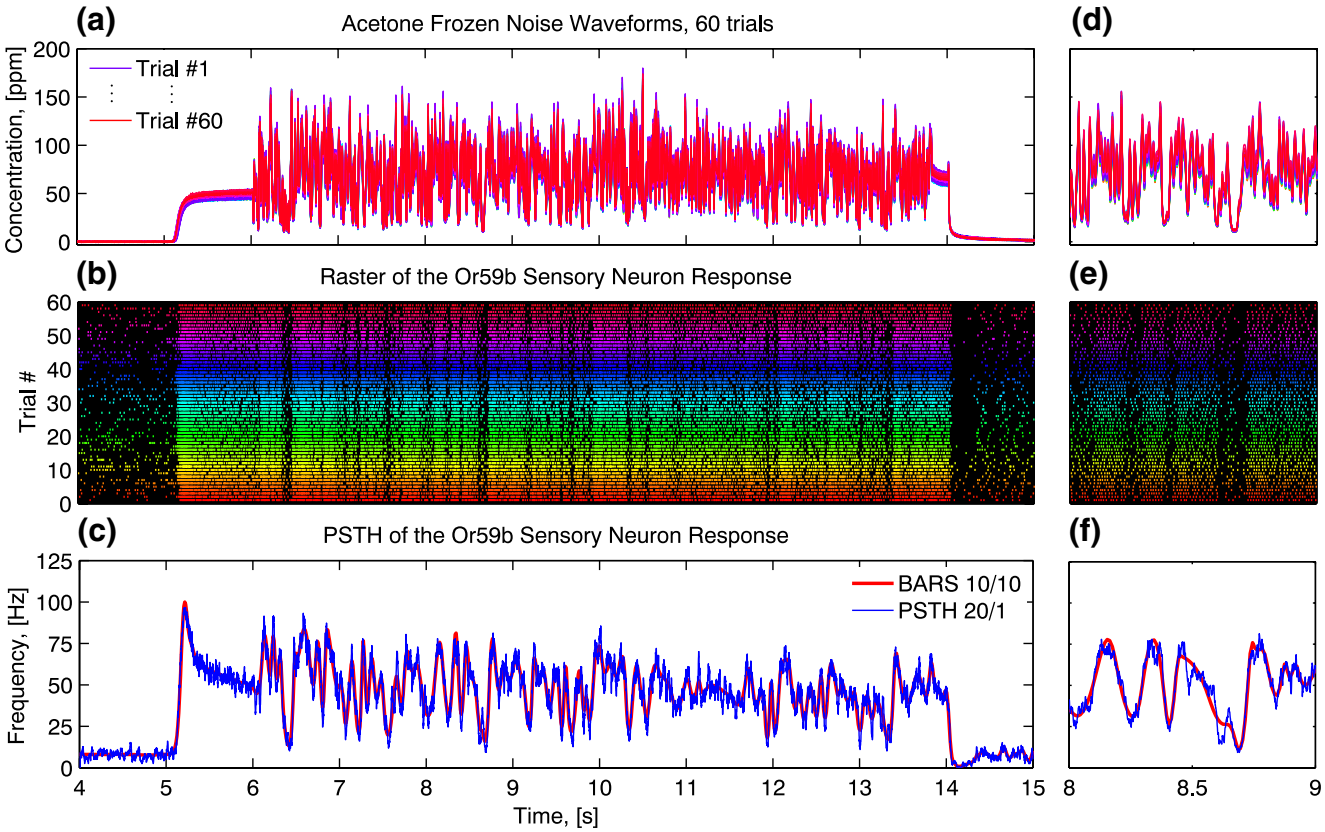

Fig. 4 The OSN response to the acetone frozen noise odor waveform. (**a**) 60 consecutive presentations of the frozen noise odor waveform. Note the remarkable reproducibility in odor delivery. (**b**) The corresponding raster of the OSN response. A different color is used for each trial # to highlight that in repeated trials the delivered odor waveforms differ by up to 1%. (**c**) The PSTH of the OSN response to the frozen noise waveform was computed using a 20 ms bin size and a 1 ms sampling interval. The BARS algorithm applied to a PSTH with non-overlapping 10 ms bins provides an additional estimate of the neural response. (**d**)–(**f**) A one-second-long window from (**a**)–(**c**)

output and the predicted output for each model. The cross-validation in Fig. 5(c) shows that the 1D LNP cascade model can closely predict the PSTH of the neuron response.

4.3.3 The 2D LNP model of the Or59b OSN

We have identified the input/output map of the Or59b OSN in response to a white noise protocol assuming a 2D LNP model as described above. The eigenvectors corresponding to the two most significant eigenvalues of the C matrix in Eq. (5) are plotted within the linear block of the model in Fig. 6(a).

The first filter $h_1(t)$ is the eigenvector corresponding to the eigenvalue with the largest magnitude. It exhibits a monophasic pattern with a spectrum similar to that of a low-pass filter with a -3 dB cut-off frequency of about 20 Hz (Fig. 6(b)). The second filter $h_2(t)$ exhibits a biphasic pattern with positive and negative peaks at around 75 ms and 150 ms. Its spectrum is similar to that of a band-pass filter with a -3 dB cut-off frequency of 1 Hz (low) and 12 Hz (high) (Fig. 6(b)).

The nonlinear block of the model was estimated by matching the output of the two filters to the PSTH using the 2D ridge estimator (see Section 2.3 in the Supplemental material). The resulting nonlinearity is plotted in Fig. 6(a).

The cross-validation in Fig. 6(c) shows that the 2D LNP cascade model can closely predict the PSTH of the neuron response. The black trace depicts the PSTH, whereas the red trace corresponds to the model prediction in response to a white noise input stimulus with mean $\mu = 68$ ppm and variance $\sigma^2 = 25^2$.

4.4 Concluding remarks

The 1D and 2D phenomenological OSN models described in the previous sections were independently derived. A simple comparison of the predictive power of these two models is given in Section 4.4.1.

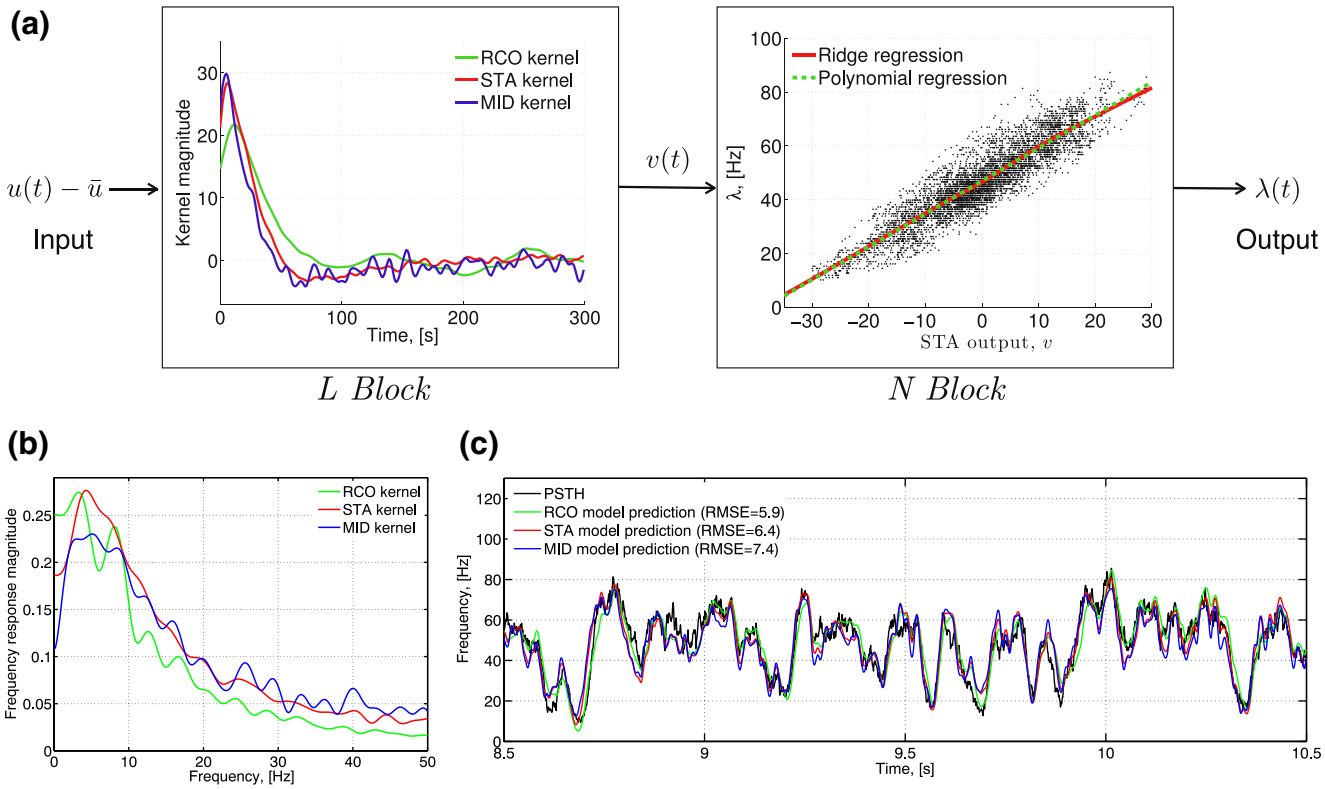

Fig. 5 The one-dimensional LNP cascade model of the Or59b OSN. (**a**) The block diagram of the 1D LNP cascade model. The model input $u(t) - \bar{u}$ is filtered with a kernel $h(t)$ and then passed through a static nonlinearity. The model output $\lambda(t)$ is the random intensity rate of the spike train. OSN kernels were estimated using the RCO, STA, and MID methods. The static nonlinearity was estimated by matching the output of the linear block to the OSN PSTH. In this example, both ridge regression and polynomial regression methods were used to obtain the nonlinearity for the STA kernel. (**b**) Spectral estimates of the RCO, STA and MID kernels. (**c**) The cross-validation shows that the 1D LNP cascade model can closely predict the OSN PSTH in response to novel noise stimuli with the same mean $\mu = 66$ ppm and contrast $\sigma/\mu = 0.38$. The RMSE between the PSTH (*black*) and the estimated random intensity rate of the spike train is 5.9 Hz, 6.4 Hz, and 7.4 Hz for the RCO (*green*), STA (*red*), and MID (*blue*) kernels, respectively

The previous neuron identification results were obtained by assuming fixed stimulus statistics as both the mean and the variance of the white noise odor waveforms were kept constant. In Section 4.4.2 we shall briefly highlight the dependence of the system identification models on the input statistics. As we shall see, the phenomenological models of OSNs derived here are strongly dependent on the first moment of the white noise odor stimuli. They are, however, largely invariant with respect to stimulus contrast changes.

4.4.1 Comparison of the 1D and 2D LNP models

How do the 1D and 2D LNP cascade models discussed in the previous sections compare? That is, how do the models depicted in Figs. 5 and 6 compare? A geometric interpretation is desirable because of its intuitive appeal. Can the two models be compared in the same space say by constructing a common 2D or 3D space? We opt here for the latter.

Because the two phenomenological models differ in their dimensionality, we shall first map the nonlinearity of the 1D model into a two-dimensional space. This is readily possible because the nonlinearity in Fig. 5(a) is parametrized by the average value of the input stimulus $\bar{u}$. In Fig. 7(a) the one-dimensional nonlinearity is shown as a function of the STA filter output and the average of the input stimulus $\bar{u}$. The STA filter used in this representation was identified for the medium stimulus mean value $\mu = 45$ ppm.

Thus, the 1D system identification model of the OSN can be interpreted as being two-dimensional. The STA filter output and an average of the stimulus are the inputs to a two-dimensional nonlinearity as shown in Fig. 7(a). Clearly, a reinterpretation of the 1D system identification model of Fig. 5(a) for different average stimulus values $\bar{u}$ leads to a straightforward comparison with the 2D LNP model in Fig. 6(a).

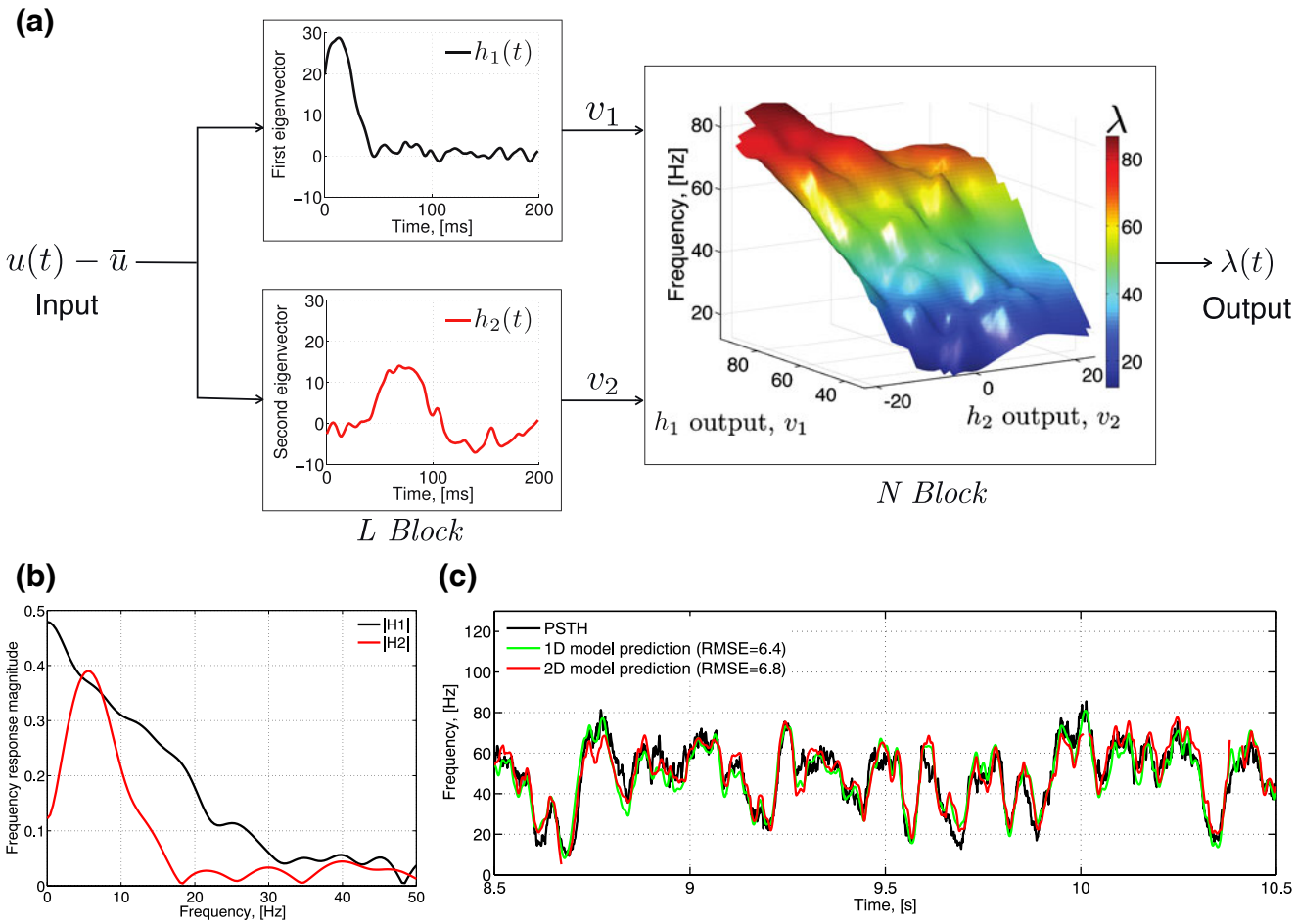

Fig. 6 The two-dimensional LNP cascade model of the Or59b OSN. (**a**) The block diagram of the 2D LNP cascade model. The model input $u(t) - \bar{u}$ is filtered with two kernels h_1 and h_2 and the filter outputs v_1 and v_2 are passed through a static 2D nonlinearity. The model output $\lambda(t)$ is the random intensity rate of the spike train. The OSN kernels h_1 and h_2 were estimated using the STC method. (**b**) The Fourier transform of the filters (H_1 and H_2) shows that h_1 is a low-pass filter with a cutoff frequency of about 20 Hz and h_2 is a band-pass filter with cutoff frequencies at roughly 1 Hz and 12 Hz. (**c**) The cross-validation shows that the 2D LNP cascade model can closely predict the OSN PSTH in response to novel noise stimuli with the same mean $\mu = 68$ ppm and contrast $\sigma/\mu = 0.38$. The RMSE between the PSTH (*black*) and the estimated random intensity rate of the spike train (*red*) is 6.8 Hz. For comparison, the 1D LNP model prediction is shown in green. The RMSE between the PSTH (*black*) and the estimated random intensity rate of the spike train (*red*) is 6.4 Hz

A comparison of the predictive power of the 1D and 2D models is shown in Fig. 6(c). The RMSE between the OSN PSTH and the response of the 1D and 2D models are 6.4 and 6.8 Hz, respectively. Thus the 2D model performs on par with the 1D LNP model with a RMSE increase of about 5 to 10%.

4.4.2 Dependence on the stimulus statistics

In Fig. 7 we also demonstrate the dependence of the system identification results on the statistics of the input stimuli. In Fig. 7(a) we plot the one-dimensional nonlinearity for three different values of $\mu = \bar{u}$, the first moment of u, while the stimulus contrast σ/μ is essentially kept constant (in what follows μ and $\bar{u}$ are interchangeably used). Note that the slope of the nonlinearity varies with different values of $\bar{u}$. While we depict the nonlinearity for only three values of $\bar{u}$ (23, 45 and 102 ppm), it is clear that a continuum of such nonlinearities exists for practically the same values of the contrast σ/μ (0.33, 0.34, 0.35). Thus, assuming that the contrast is kept fixed, the system identification results for the 1D LNP model depend upon the average value of the input stimulus.

A similar dependence is observed in Fig. 7(b) for the 2D LNP model. The filter pair (h_1, h_2) was evaluated

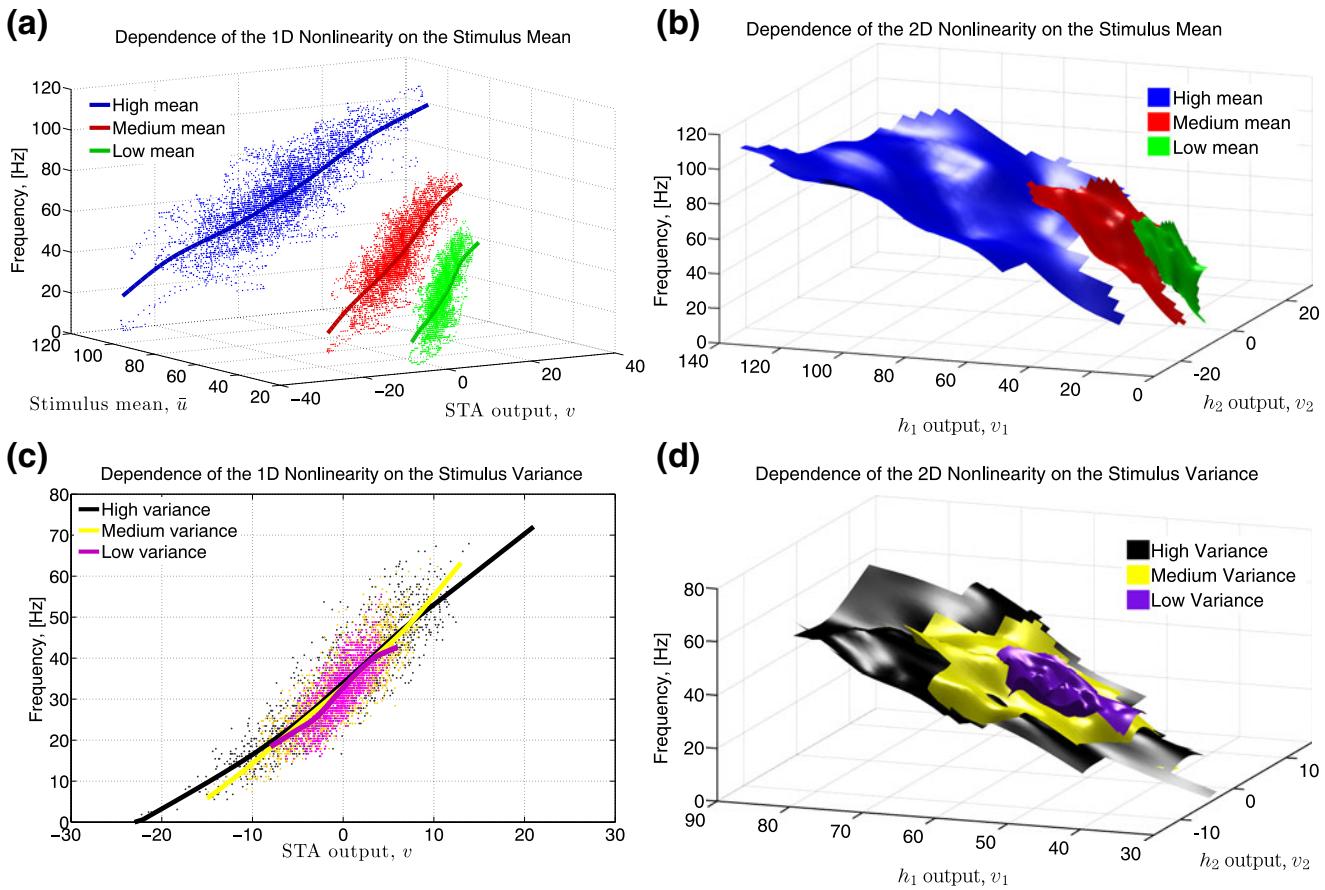

Fig. 7 The dependence of white noise system identification results on the statistics of input stimuli. (**a**) The nonlinearity of the 1D LNP model is shown as a function of the output of the STA filter for three different values of $\bar{u}$: 23 ppm (*green*), 45 ppm (*red*) and 102 ppm (*blue*) with the corresponding contrast σ/μ values of 0.33, 0.34 and 0.35. The slope of the nonlinearity decreases with $\bar{u}$. (**b**) The nonlinearity of the 2D LNP model also changes with $\bar{u}$. Note the difference in (overlapping) nonlinearities for the three different values of $\bar{u}$ (same as in (**a**)). (**c**), (**d**) For a constant average value μ of the stimulus, the stimulus contrast σ/μ does not qualitatively affect the estimated 1D and 2D nonlinearities. In this example, $(\mu, \sigma/\mu)$ takes the values (58, 0.1) or low, (62, 0.19) or medium, and (67, 0.26) or high contrast, respectively. In both the 1D and the 2D case, the nonlinearities simply cover a larger filter output space with the increasing contrast

for the medium mean stimulus value using the STC method. Here we depict 2D nonlinearities for the same 3 value pairs of $(\mu, \sigma/\mu)$ as in Fig. 7(a). A clear gap is observed between all 3 surfaces even though their domains overlap. A representation of the footprint for a range of mean concentration values is shown in Fig. S4 in the Supplemental material.

If however the contrast of the input stimulus is varied while the mean is kept constant, no significant change in the system identification results is observed. In both cases, a higher contrast simply allows us to cover more of the space at the output of the filter (Fig. 7(c) and (d)). The mean value and contrast pair of the stimuli employed in Fig. 7(c) and (d) are, respectively, $(58, 0.1)$, $(62, 0.19)$ and $(67, 0.26)$.

To summarize, the system identification results for both the 1D and the 2D LNP model vary significantly with the statistics (specifically the first order moment) of the test odor stimuli. In a cross-validation, both models perform well when predicting response to novel stimuli with the same statistics as the test stimuli. However, the cross-validation fails when the input stimuli have a mean that differs from that of the test stimuli. Because this phenomenon is observed for both the 1D and the 2D LNP model, it points to adaptive coding properties of the OSN. Thus, the processing and representation of the olfactory information appears to be signal-dependent.

5 Discussion

In Sections 3 and 4 we presented an empirical and principled system identification methods, respectively,

and employed them to construct phenomenological models of OSNs. The basic elements of these models, the processing filters and the nonlinearities, are strongly dependent on the signal shapes and moment statistics. Consequently, the space of input stimuli was appropriately restricted in these sections. How do these models compare across different input stimulus shapes and statistics?

5.1 Exploring the space of odor stimuli

In Section 3 we characterized Or59b OSNs in response to odor waveforms that explore the input stimulus space along the odor concentration and concentration gradient (rate of change). Each triangle odor waveform sweeps a given concentration range with a unique rising/falling gradient value set, and this allows for a "more uniform" sampling of the input space.

Since the 2D Encoding Manifold from the triangle protocol and the 2D nonlinearity from the white noise protocol are plotted in different coordinate systems, a change of coordinates is required for a meaningful comparison of the two. In order to compare the empirical model in Section 3 with the principled models in Section 4, we shall assume that the shape of the STC kernels is preserved for any arbitrary odor signal even though the slope of the nonlinearity is input modulated. Therefore, triangle odor waveforms at the input of the 2D LNP cascade model are assumed to be filtered by the two STC kernels (h_1, h_2) derived using the white noise protocol for a stimulus with an average value in the medium range (see Fig. 7). We undertook a coordinate transformation of the axes in Fig. 3(c) and mapped the coordinates *(amplitude, gradient)* to (h_1, h_2) (see Fig. S7 in the Supplemental material). A highly intuitive, geometric comparison is now possible.

The resulting input/output relationship leads us to the (black) 2D Encoding Manifold in Fig. 8(a). As before, the 2D Encoding Manifold was obtained by applying the ridge-regression method described in Section 2.2 of the Supplemental material. The shape of the nonlinearity indicates that the system encodes the output v_2 of the h_2 filter "more strongly" than the output v_1 of the h_1 filter. The input space has a large footprint as shown in Fig. 8(b) (also in black).

Overlaid on the transformed 2D Encoding manifold in Fig. 8(a) are three LNP nonlinearities derived for three mean odor concentration values. The nonlinearities were obtained using the white noise odor protocol with the contrast of the odor waveforms kept constant. Green, red and blue patches depict the 2D nonlinearities corresponding to the (mean, contrast) pairs (23, 0.33) or low, (45, 0.34) or medium, and (102, 0.35) or high mean, respectively. Clearly, the white noise protocol only explores a small subset of possible encodings of the input space.

Finally, in order to test how well the 2D Encoding Manifold predicts the response of an OSN under different stimulus conditions, we first applied triangle odor waveforms at the input of the identified linear system (h_1, h_2) described above. The outputs (v_1, v_2) of the filters in response to the three arbitrary triangular odor waveforms (similar to the waveforms in Fig. 3(b)) are mapped into the green, red and blue traces on the black 2D Encoding Manifold in Fig. 9(a). The same three green, red and blue (time) traces are cross-validated with the black OSN PSTH in Fig. 9(b)–(d).

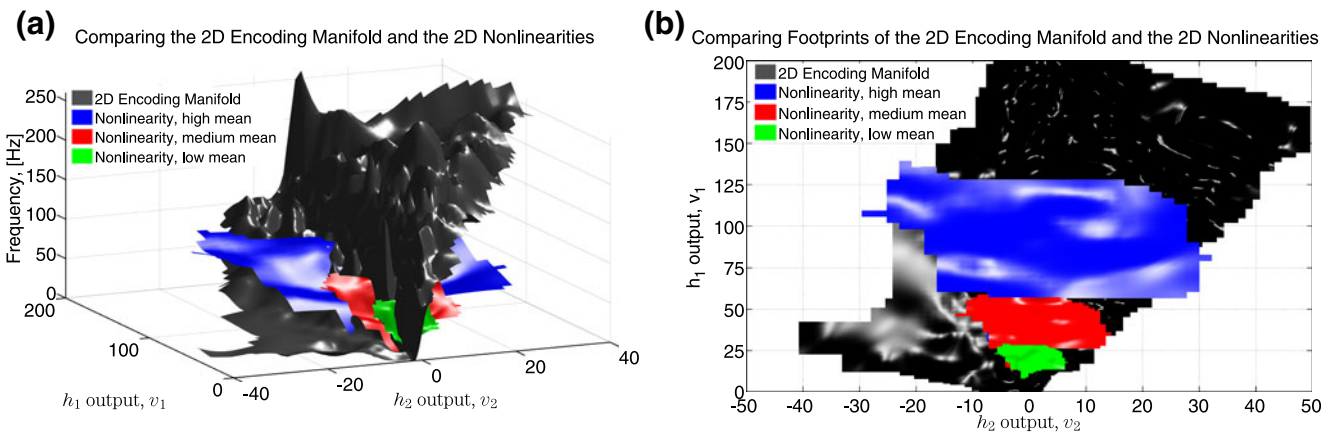

Fig. 8 Comparison of the 2D Encoding Manifold and the 2D nonlinearities for fixed contrast. (**a**) The nonlinearities of the 2D LNP model for the white noise odor protocol (*green*, *red* and *blue*; same as in Fig. 7(b)) and the 2D Encoding Manifold of the triangle odor waveform protocol (*black*). (**b**) Footprint of the nonlinearities of the 2D LNP model (*green*, *red* and *blue*) and the 2D Encoding Manifold (*black*)

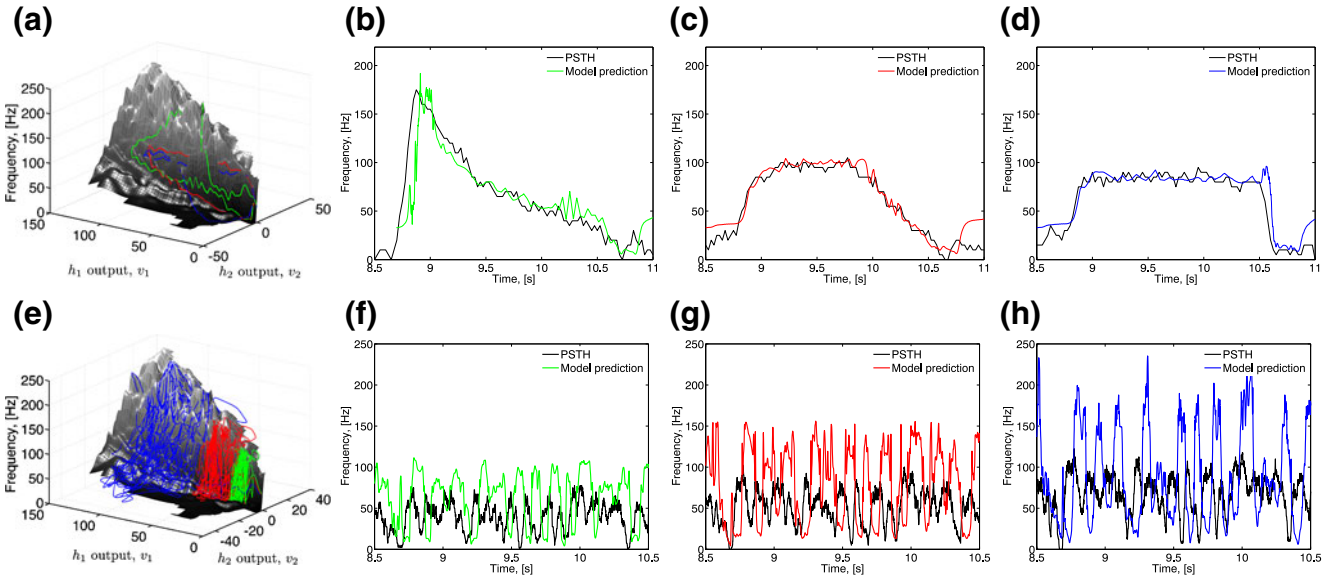

Fig. 9 Using the 2D Encoding Manifold to predict the response of an OSN. (**a**) Three arbitrary triangle odor waveforms (*green*, *red* and *blue*) mapped onto the 2D Encoding Manifold (*black*). Sample odor waveforms are shown in Fig. 3(b). (**b**)–(**d**) Cross-validation using the three arbitrary triangle waveforms (*green*, *red* and *blue*) and the OSN PSTH (*black*). (**e**) White noise odor waveforms mapped onto the 2D Encoding Manifold. The white noise odor waveforms have the mean contrast pairs (58, 0.1) in *green*, (62, 0.19) in *red*, and (67, 0.26) in *blue*, respectively. (**f**)–(**h**) Cross-validation of three white noise odor waveforms (*green*, *red* and *blue*) and the corresponding OSN PSTH (*black*)

Three white noise odor waveforms (shown in green, red and blue) at the input of the filter pair (h_1, h_2) are first transformed into the outputs (v_1, v_2) and then mapped as traces onto the black 2D Encoding Manifold in Fig. 9(e). Figure 9(f)–(h) show the time traces in the same color code together with the black OSN PSTH of the recorded spike trains. The model neuron consisting of the (h_1, h_2) filter pair and the 2D Encoding Manifold cascade exhibits a strong predictive capability for triangle odor waveforms. It has, however, a limited utility for predicting the response to white noise odor stimuli.

Therefore, the identification of OSNs using white noise odor stimuli leads to a phenomenological description that is quite different from the empirical model developed based on triangle odor waveforms. Both models have their respective domains of validity, however. The empirical model can predict the response to slowly varying triangle odor waveforms (but not white noise) whereas the 2D LNP model can predict the response to white noise (but not to non-stationary triangle stimuli).

5.2 Final remarks

Fundamentally, system identification deals with the construction of mathematical models of dynamical systems based on the response of these systems to test stimuli. As such, system identification requires a precise control of the test stimuli and accurate measurements of both the input and the output of the system.

In olfaction, system identification of sensory neurons has had limited success primarily because no adequate measurement of the concentration of time-varying odor waveforms was available. To our knowledge, the only exceptions are (Justus et al. 2005; French and Meisner 2007; Schuckel et al. 2008). However, in these works the system response is evaluated using electroantennograms that do not provide precise spike-time recordings.

In this study we investigated the problem of system identification in olfactory sensory neurons of the fruit fly using a novel odor delivery and acquisition system that allowed for a precise control and measurement of odor concentration on a millisecond time scale. We recorded the extracellular activity of olfactory sensory neurons in response to various test stimuli and investigated the empirical as well as several classes of mathematical models of OSNs. Two methods of system identification were employed. They differ both in terms of the stimuli used and the methodology of identifying the OSNs.

First, we used triangle odor stimuli and an empirical method of system identification. The resulting empirical model encodes both the odor concentration and

the odor concentration gradient with a 2D Encoding Manifold. The form of the manifold can be estimated from the spike time recordings using a ridge estimator. The empirical model already demonstrates that the first layer of the olfactory system is, computationally, more sophisticated than previously thought. Consequently, we cannot disregard the fine structure of the OSN response at the periphery of the olfactory system and hope to understand odor signal processing in the higher brain centers. More fundamental research on the early olfactory system is needed in order to understand the odor representation and processing by OSNs.

Second, we employed white noise odor stimuli and applied a principled approach to identify OSNs. The 1D and 2D LNP cascade phenomenological models of OSNs were investigated. For white noise odor waveforms as input stimuli we built a 1D LNP cascade model of the sensory neuron for a fixed mean odor concentration and fixed contrast. In this model a single linear filter is followed by a 1D nonlinearity. The instantaneous response of the neuron was determined by passing a mean-zero signal through the linear filter and then through the nonlinearity. Although such a model can predict the response to stimuli with the same mean and contrast, it fails to do so when the mean concentration is altered while the contrast remains unchanged. For a given fixed contrast we demonstrated that the nonlinear block of the LNP cascade model changes with the mean concentration of the stimulus. This points to the fact that the OSN model should be at least two-dimensional.

We then investigated a 2D LNP cascade model in which the stimulus is passed through two parallel linear filters and the output of the filters is fed into a two-dimensional nonlinearity. By comparing the geometry of the input/output map associated with the nonlinear block, we showed that a standard white noise analysis only provides a partial view of both the footprint and the transformation of the 2D system. The footprint can easily be enlarged through the delivery of triangular odor waveforms, or more generally, time-dependent waveforms. Exploring other characteristics of the transformation will require experimentation with additional odor waveforms.

Finally, we evaluated a 2D model of Or59b OSNs consisting of a pair of filters in cascade with the 2D Encoding Manifold using both stationary white noise and non-stationary triangle odor stimuli. This model exhibits a strong predictive capability for triangle odor waveforms. It has, however, a limited capability for predicting the response to white noise odor stimuli. It suggests an adaptive neural encoding model for Or59b OSNs with a nonlinearity that depends, for white noise stimuli with fixed contrast, on the mean odor waveform. Contrast independence for fixed mean concentration values was a surprising find.

System identification of OSNs using triangular odor waveforms offers a number of advantages to the systems neuroscientist because of the fine level of parametrization of the input space. We note that extensive electrophysiology with the same level of stimulus control has been performed by other researchers on various neurons and in other organisms. A few examples include the thermo-receptor neurons in Loftus (1969), hygro-receptor neurons in Tichy (2003), and the olfactory sensory neurons in Hinterwirth et al. (2004) and Tichy et al. (2005). In these studies the output of sensory neurons were parametrized using two input signal components: the input amplitude and its rate of change. Furthermore, even in response to white noise waveforms, two-dimensional models include two linear kernels that extract the input amplitude and its rate of change (Brenner et al. 2000; Slee et al. 2005). In this light, our work on the system identification of Or59b OSNs bridges the gap between the high degree of control of the concentration and the concentration gradient of odor waveforms and the inference that can be made about the actual stimulus encoding of these neurons.

The RCO, STA and MID kernels employed, while closely related, seem to be stimulus-dependent and might change depending on the state of the system. Note that, e.g., the STA kernel has a fundamental limitation in that it depends on the statistics of the stimulus (Paninski 2003). Furthermore, a memoryless Poisson process was assumed that does not really capture the temporal statistics of neural spike trains (Berry and Meister 1998; Keat et al. 2001; Reich et al. 1998; Aguera y Arcas and Fairhall 2003). In addition, the spike history dependence can bias the estimation of linear filters (Berry and Meister 1998; Paninski et al. 2003; Paninski 2003; Aguera y Arcas and Fairhall 2003).

Finally, the results presented here are limited to the class of Or59b OSNs. If and whether the insights provided here arise in OSNs expressing other receptor types will be explored elsewhere.

Acknowledgements The work presented here was supported by NIH under grant number R01DC008701-01 and was conducted in the Axel laboratory at Columbia University. The authors would like to thank Dr. Richard Axel for insightful discussions and for his outstanding support. The authors would like to also thank the reviewers for their suggestions for improving the presentation of the paper.

References

Aertsen, A. M. H. J., & Johannesma, P. I. M. (1981). The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biological Cybernetics, 42*, 13343.

Aguera y Arcas, B., & Fairhall, A. (2003). What causes a neuron to spike? *Neural Computation, 15*, 1789–1807.

Bau, J., Justus, K., & Carde, R. (2002) Antennal resolution of pulsed pheromone plumes in three moth species. *Journal of Insect Physiology, 48*(4), 433–442.

Berry, M. J., & Meister, M. (1998). Refractoriness and neural precision. *Journal of Neuroscience, 18*, 2200–2211.

Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Statistical properties of spike trains: Universal and stimulus-dependent aspects. *Neuron, 26*(3), 695–702.

Bussgang, J. (1952). *Crosscorrelation functions of amplitude-distorted Gaussian signals* (Vol. 216, p. 3). Technical Report, Research Laboratory Of Electronics, MIT.

Clyne, P., Grant, A., O'Connell, R., & Carlson, J. (1997). Odorant response of individual sensilla on the *Drosophila* antenna. *Invertebrate Neuroscience, 3*(2), 127–135.

de Bruyne, M., Foster, K., & Carlson, J. (2001). Odor coding in the drosophila antenna. *Neuron, 30*(2), 537–552.

Dougherty, D. P., Wright, G. A., & Yew, A. C. (2005). Computational model of the cA.M.P.-mediated sensory response and calcium-dependent adaptation in vertebrate olfactory receptor neurons. *PNAS, 102*, 2415.

Fairhall, A. L., Burlingame, C. A., Narasimhan, R., Harris, R. A., Puchalla, J. L., & Berry, M. J. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *Journal of Neurophysiology, 96*(5), 2724–2738.

French, A. S., & Meisner, S. (2007). A new method for wide frequency range dynamic olfactory stimulation and characterization. *Chemical Senses, 32*(7), 681–688.

Geffen, M. N., Broome, B. M., Laurent, G., & Meister, M. (2009). Neural encoding of rapidly fluctuating odors. *Neuron, 61*(4), 570–586.

Gomez, G., Voigt, R., & Atema, J. (1999). Temporal resolution in olfaction III: Flicker fusion and concentration-dependent synchronization with stimulus pulse trains of antennular chemoreceptor cells in the American lobster. *Journal of Comparative Physiology A, 185*(5), 427–436.

Gu, Y., Lucas, P., & Rospars, J.-P. (2009). Computational model of the insect pheromone transduction cascade. *PLoS Computers in Biology, 5*(3), e100321.

Halnes, G., Ulfhielm, E., Ljunggren, E. E., Kotaleski, J. H., & Rospars, J.-P. (2009). Modelling and sensitivity analysis of the reactions involving receptor, G-protein and effector in vertebrate olfactory receptor neurons. *Journal of Computational Neuroscience, 27*, 471.

Hinterwirth, A., Zeiner, R., & Tichy, H. (2004). Olfactory receptor cells on the cockroach antennae: Responses to the direction and rate of change in food odour concentration. *European Journal of Neuroscience, 19*, 3389–3392.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation. *Journal of Physiology, 117*, 500–557.

Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and hammerstein cascade models. *Biological Cybernetics 55*(2–3), 135–144.

Justus, K., Carde, R., & French, A. S. (2005). Dynamic properties of antennal responses to pheromone in two moth species. *Journal of Neurophysiology, 93*, 2233–2239.

Kaissling, K.-L. (2009). Olfactory perireceptor and receptor events in moths: A kinetic model revised. *Journal of Comparative Physiology, 195*, 895.

Kass, R. E., Ventura, V., & Cai, C. (2003). Statistical smoothing of neuronal data. *Network: Computation in Neural Systems, 14*, 5–15.

Kass, R. E., Ventura, V., & Brown, E. N. (2005) Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology, 94*, 8–25.

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron, 30*(3), 803–817.

Lemon, W., & Getz, W. (1997). Temporal resolution of general odor pulses by olfactory sensory neurons in American cockroaches. *Journal of Experimental Biology, 200*(Pt 12), 1809–1819.

Lindemann, B. (2001). Predicted profiles of ion concentrations in olfactory cilia in the steady state *Biophysical Journal, 80*, 1712.

Loftus, R. (1969). Differential thermal components in the response of the antennal cold receptor of Periplaneta americana to slowly changing temperature. *Journal of Comparative Physiology A, 63*(4), 415–433.

Marmarelis, V. (2004). *Nonlinear dynamic modeling of physiological systems*. Wiley-IEEE Press.

Marmarelis, P. Z., & Naka, K. (1972). White-noise analysis of a neuron chain: An application of the Wiener theory. *Science, 175*(27), 1276–1278

Paninski, L. (2003). Convergence properties of three spike-triggered analysis techniques. *Network: Computation in Neural Systems, 14*(3), 437–464.

Paninski, L., Lau, B., & Reyes, A. (2003) Noise-driven adaptation: *In vitro* and mathematical analysis. *Neurocomputing, 52*, 877–883.

Pillow, J. (2007). Likelihood-based approaches to modeling the neural code. In: *Bayesian brain: Probabilistic approaches to neural coding* (chap. 3). MIT Press.

Pillow, J., & Simoncelli, E. (2006). Dimensionality reduction in neural models: An information-theoretic generalization of the spike-triggered average and covariance analysis. *Journal of Vision, 6*, 414–428.

Reich, D., Victor, J., & Knight, B. (1998). The power ratio and the interval map: Spiking models and extracellular recordings. *Journal of Neuroscience, 18*, 10,090–10,104.

Reidl, J., Borowski, P., Sensse, A., Starke, J., Zapotocky, M., & Eiswirth, M. (2006). Model of calcium oscillations due to negative feedback in olfactory cilia. *Biophysical Journal, 90*, 1147.

Rospars, J.-P., Lansky, P., Duchamp, A., & Duchamp-Viret, P. (2003). Relation between stimulus and response in frog olfactory receptor neurons *in vivo*. *European Journal of Neuroscience, 18*, 1135.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience, 9*(11), 1421–1431.

Schuckel, J., Meisner, S., Torkkeli, P., & French, A. S. (2008). Dynamic properties of *Drosophila* olfactory electroantennograms. *Journal of Comparative Physiology A, 194*, 483–489.

Schwartz, O., Pillow, J. W., Rust, N. C., & Simoncelli, E. P. (2006). Spike-triggered neural characterization. *Journal of Vision 6*(4), 484–507.

Sharpee, T., Rust, N. C., & Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Computation 16*(2), 223–250.

Slee, S., Higgs, M., Fairhall, A., & Spain, W. (2005). Two-dimensional time coding in the auditory brainstem. *Journal of Neuroscience, 25*(43), 9978–9988.

Steveninck, R. V., & Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. *Proceedings of the Royal Society of London Series B, 24*, 379–414.

Tichy, H. (2003). Low rates of change enhance effect of humidity on the activity of insect hygroreceptors. *Journal of Comparative Physiology A, 189*, 175–179.

Tichy, H., Hinterwirth, A., & Gingl, E. (2005). Olfactory receptors on the cockroach antenna signal odour ON and odour OFF by excitation. *European Journal of Neuroscience, 22*, 3147–3160.

Victor, J. D. (2005). Analyzing receptive fields, classification images and functional images: Challenges with opportunities for synergy. *Nature Neuroscience, 8*, 1651–1656.

Victor, J., & Shapley, R. (1980). A. method of nonlinear analysis in the frequency domain. *Biophysical Journal, 29*, 459–484.

Wiener, N. (1958). *Nonlinear problems in random theory*. MIT Press.

Wu, M., David, S., & Gallant, J. (2006). Complete functional characterization of sensory neurons by system identification. *Annual Review of Neuroscience, 29*, 477–505.