Odorant Concentration Dependence in Electroolfactograms Recorded From the Human Olfactory Epithelium

Hadas Lapid,1,2 Han-Seok Seo,3 Benno Schuster,3 Elad Schneidman,1 Yehudah Roth,4 David Harel,2 Noam Sobel,1,* and Thomas Hummel1,4*

1Departments of Neurobiology and 2Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel; 3Smell and Taste Clinic, University of Dresden Medical School, Dresden, Germany; and 4Department of Otolaryngology-Head and Neck Surgery, Edith Wolfson Medical Center, Tel-Aviv University Sackler School of Medicine, Holon, Israel

Submitted 15 December 2009; accepted in final form 5 July 2009

INTRODUCTION

A unique opportunity to directly record neural activity in humans in vivo is presented by the olfactory system (Zelano and Sobel 2005). At its peripheral end, the olfactory system is equipped with bipolar sensory neurons (ORNs) that are readily accessible in the nose. Here we set out to ask whether a local field potential (LFP) measured directly from these neurons, also known as the electroolfactogram (EOG), is odor-concentration dependent.

The EOG is a slow negative voltage change recorded extracellularly from a population of neurons in response to an olfactory stimulus (Chaput 2000; Chaput and Chalansonnset 1997; Edwards et al. 1988; Ezeh et al. 1995; Furukawa et al. 1989; Mackay-Sim and Kesteven 1994; Mozell 1962; Ottoson 1955, 1958; Scott 2006; Scott et al. 1996, 1997; Thommesen and Doving 1977; Wang et al. 1993). EOGs are commonly characterized by the response amplitude that is a negative peak denoted by N1 and by the response latency that relates to the time point of the amplitude (L1).

In humans, EOGs were first measured by Osterhammel et al. (1981), and the technique was later improved by Kobal (1981). Several lines of evidence suggest that human EOGs indeed reflect the earliest level of odor coding. For example, an EOG has been measured following presentation of subthreshold odorant stimuli that were not consciously perceived (Hummel et al. 2006), and successive presentation of odorants lead to perceptual habituation before it was reflected in EOG habituation (Hummel et al. 1996). In fact, the human EOG response may be robust to the extent that evoked potentials, with many of the characteristics of the EOG recorded from the olfactory mucosa, can be recorded externally at a site close to the bridge of the nose (Wang et al. 2004).

The preceding findings can be taken as an indication to the power of the EOG method. Conversely, a minimally habituating signal that is obtained without perception may raise the concern of nonsensory (e.g., respiratory) sources of signal. Key features of sensory signals are dose dependence and identity specificity. These features have been validated in rodent and insect EOGs (Mozell 1964; Mustaparta 1971, 1990; Scott et al. 2000, 2006) and were depicted in brief within the thesis describing development of the human EOG method (Kobal 1981), but they have not been tested comprehensively in humans. Thus the aim of this study was to probe for the first of these two key features, namely dose dependence, to further validate EOG as a method to probe sensory coding in humans.

METHODS

Subjects

Sixty-two subjects enrolled in the study after providing informed written consent to procedures approved by the Ethics Committee of the Medical Faculty of the Technical University of Dresden. All subjects scored 10 or above on a 12-item forced-choice identification task of common household odors (Sniffin’ Sticks, Burghart GmbH, Wedel, Germany) (Hummel et al. 2001). Subjects had no history of neurological or sinususal disease. Endoscopy of the nose did not reveal any major abnormalities in any of the subjects. These subjects were all experienced at the method of velopharyngeal closure (Kobal 1981), a breathing technique that prevents nasal airflow during an experimental trial. Eighteen of the subjects (6 women, mean age = 23.5 yr) participated in a two session psychophysical test only. Recording EOGs was attempted in 44 subjects. The experiment concluded prematurely in 50% of the cases, primarily due to difficulties in placing the electrode at areas of functional olfactory epithelium. Thus EOG data from 22 subjects (12 women, mean age = 23.3 yr) who completed the experiment was submitted to further analysis.

Odorants

Four odorants supplied by Sigma-Aldrich (Israel) were used: isovaleric acid (CAS 504-74-2), valeric acid (CAS 109-52-4), linalool (CAS 1338-86-9), and valeric acid and linalool (12 women, mean age = 23.3 yr) was tested with two odorants, either valeric acid and linalool (n = 12) or isovaleric acid and 1-carvone (n = 10), each delivered at four concentrations diluted with warm (37°C) and humidified (80%) odorless air. In behavior, increased odorant concentration was associated with increased perceived intensity (all F > 5, all P < 0.001). In EOG, increased odorant concentration was associated with increased area under the EOG curve (all F > 8, all P < 0.001). These findings substantiate EOG as a tool for probing olfactory coding directly at the level of olfactory receptor neurons in humans.

* N. Sobel and T. Hummel contributed equally to his work.

Address for reprint requests and other correspondence: H. Lapid, Dept. of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel (E-mail: hadas.lapid@weizmann.ac.il).
(CAS 78-70-6) and l-carvone (CAS 6485-40-1). To equate the perceived intensity across odors, an independent pilot pretest was conducted [10 subjects rated each stimulus 4 times, interstimulus interval (ISI) = 30 s, total number of trials = 36, using a number-free visual-analogue scale (VAS) ranging between “extremely weak,” representing the scale mark 0, and “extremely intense,” representing the scale mark 10]. This resulted in diluting all odors outside the olfactometer to 2% with propylene glycol (Fluka, CAS 57-55-6). The average perceived intensity ratings (±SD) for the 2% dilutions were for isovaleric acid 4.57 ± 1.1, for valeric acid 3.5 ± 1.9, for linalool 4.01 ± 0.99, and for l-carvone 5.09 ± 1.1. No differences were observed between the paired odors in each experiment [iso-valeric acid and l-carvone: T(9) = 0.47, P = 0.649, valeric acid and linalool: T(9) = 0.97, P = 0.357], thus validating our pilot perceived-intensity-equation effort.

Odorant delivery

For chemical stimulation, a computer-controlled air-dilution olfactometer (OM6b, Burghart Instruments, Wedel, Germany) was used for delivering the odorants to one of the nostrils without altering mechanical or thermal conditions at the stimulated mucosa. This monomodal chemical stimulation was achieved by mixing pulses of the stimulants in a constantly flowing air stream with controlled temperature (37°C) and humidity (80% RH). The air stream was led into the nasal cavity with Teflon tubing (8 cm length, 2 mm ID).

Psychophysics experiment design

To obtain perceived intensity estimates, we conducted a psychophysical experiment. To reduce deviations from physiological conditions at the mucosa each of the odorants, diluted to 2% with polyethylene glycol outside the olfactometer, was further diluted inside the olfactometer with heated (37°C) and humidified (80%) clean air. The odorants were delivered at four dilution steps equal in final flow rate ($S_{\text{final}}$). The olfactometer dilutions were: 0, 12.5, 25, 37.5, and 50%. In other words, of the $S_{\text{final}}$ final flow, the odorant flows were: 0, 1, 2, 3, and 4, for each of the dilution steps respectively (thus the clean air dilution was 8, 7, 6, and 5, respectively). Events were presented in random order with minimal number of four repeats per event. A minimal ISI of 30 s was kept with stimulus duration of 500 ms starting at $t = 0.5s$. After each stimulus, subjects were asked to evaluate the perceived intensity of the stimulus on a number-free visual analogue scale ranging between extremely weak (representing the scale mark 0) and extremely strong (representing the scale mark of 10).

EOG recording apparatus

EOGs were recorded by means of Ag/AgCl electrode coated with Teflon tubing (0.8 mm OD) filled with Ringer-agar (1%). The Teflon sheath extended 0.5 mm beyond the wire tip, such that conduction was via the Ringer-agar bridge. The electrodes were freshly coated with Teflon tubing filled with the Ringer-agar solution before each experiment ($\leq 3$ h before recording). Electrode impedance was $\leq 10 \Omega$, sampling rate was 62.5 Hz; amplifier impedance was 10 M$\Omega$, with a band-pass of 0.1–15 Hz. For reference, an Ag/AgCl electrode was placed on the contralateral bridge of the nose. Another reference electrode was placed on the earlobe, two ground electrodes were placed on the subject’s mastoids, and another electrode was placed above one of the eyebrows for eye-blink tracking (Fig. 1A). The recording electrode was inserted into the nasal cavity guided by a rigid endoscope and placed on the mid or ventral part/insertion of the middle turbinate. After the electrode was positioned, it was stabilized by means of adjustable clips on a frame similar to lensless glasses (Hummel et al. 1996). Final adjustments of the tip of the electrode were performed by the experimenter under endoscopic control. After an EOG had been established in response to either of the tested odors at their highest test concentrations, the session began. During stimulation onset subjects were asked to hold their breath.

EOG experiment design

To limit subject discomfort, experimental sessions were limited to $\sim 1$ h in duration. This allowed us to test one of the odorant pairs in each subject, either iso-valeric acid/l-carvone ($n = 10$) or valeric acid/linalool ($n = 12$). The EOG experiment was identical in design to the psychophysics experiment described in the preceding text.

Analysis

PSYCHOPHYSICS. Perceived intensity estimates. To test whether the different concentrations, we used were perceived at different intensities we first normalized the subject’s average estimates for each odorant with its respective dilutions (Eq. 1, Fig. 4), and then performed a repeated-measures ANOVA followed by planned one-tailed tests with Benjamini-Hochberg correction for multiple comparisons. These differences are summarized in Table 3.

Equation 1: Normalization of intensity estimates

$$I_{i,C}^n = \frac{I_{i,C}^{\text{raw}} - \text{Min}(I_i)}{\text{Max}(I_i) - \text{Min}(I_i)}$$

To eliminate scaling variability across subjects, subject’s average intensity estimates were normalized across concentrations (0, 12.5, 25, 37.5, and 50%) for each odorant separately. $I_{i,C}^n$ is the normalized intensity estimate of an odorant at concentration $C$ in subject $i$. $\text{Min}(I_i)$ is the minimal intensity estimate across the odorant’s concentrations including blank in subject $i$. $\text{Max}(I_i)$ is the maximal intensity estimate across the odorant’s concentrations including blank.
TABLE 1. Variance within and across subjects

<table>
<thead>
<tr>
<th>Odorant</th>
<th>$\Delta V_{\text{within}}$, mV</th>
<th>$\Delta V_{\text{across}}$, mV</th>
<th>df</th>
<th>$T$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso-valeric acid</td>
<td>0.0277</td>
<td>0.0563</td>
<td>(18,2)</td>
<td>4.564</td>
<td>0.0002</td>
</tr>
<tr>
<td>l-carvone</td>
<td>0.0363</td>
<td>0.0554</td>
<td>(22,2)</td>
<td>3.702</td>
<td>0.002</td>
</tr>
<tr>
<td>Valeric acid</td>
<td>0.0212</td>
<td>0.0877</td>
<td></td>
<td>5.577</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Linalool</td>
<td>0.0230</td>
<td>0.0504</td>
<td></td>
<td>5.429</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Paired $t$-test between electroolfactographic (EOG) distance within and across subjects (Eqs. 2 and 3, respectively). Each of the four odorants was tested separately. $N$ (iso-valeric acid) = $N$ (l-carvone) = 10, $N$ (valeric acid) = $N$ (linalool) = 12).

EOG

TRIAL REJECTION CRITERIA. Eye blinks were tracked throughout the experimental sessions. Events concurrent with eye blinks or other motion artifacts were rejected. These consisted of 10.7 ± 1.95% of trials per subject. Also trials that did not elicit any response (null trials) were rejected. These consisted of 6.3 ± 1.3% of trials per subject. Critically, trial rejection was conducted as an initial stage before any detailed analysis.

Assessing EOG variation within and between subjects

To quantify EOG variation within subjects, we calculated the average distance between the evoked responses to the same odorants within subjects. Specifically, we averaged over the integrated differences between repeated measures of each odorant in a time window of a typical evoked response ($t_0 = 0.5\;s$ to $t_r = 2.5\;s$, Eq. 2).

Equation 2: EOG distance within subjects

$$\Delta V_{\text{within}} = \frac{1}{\Delta t \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{2s} | V_{i,j}(t) - V_{i,m}(t) |$$

$\Delta V_{\text{within}}$ is the average EOG distance between responses evoked by an odorant within subjects. $V_{i,j}(t)$ is the $j$th response in subject $i$ at time $t$. $n_i$ is the total number of successful recordings in subject $i$. $m$ index iterates over the remaining $n_i - 1$ responses evoked by the same odorant in the $i$th subject but the $j$th response. $N$ is the number of subjects. $\Delta t = 2\;s$, the time window of the evoked response. $t$ integrates over all time steps in that window with time bins of 16 ms (62.5 Hz).

To assess EOG variation between subjects, we calculated the average distance between the responses evoked by the same odorant in different subjects. Here the integrated differences were calculated between each of the repeated measures in a specific subject and the remaining responses in the other subjects (Eq. 3).

Equation 3: EOG distance between subjects

$$\Delta V_{\text{across}} = \frac{1}{\Delta t \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{2s} | V_{i,m}(t) - V_{i,j}(t) |$$

$\Delta V_{\text{across}}$ is the average EOG distance between responses evoked by an odorant in different subjects. $V_{i,m}(t)$ is the $m$th response of subject $i$ at time $t$. $V_{i,j}(t)$ is the $j$th response of subject $j$ at time $t$. $n_i$ and $n_j$ are the number of EOGs evoked by the tested odorant in subjects $i$ and $j$, respectively.

A paired $t$-test was then performed between EOG distance within subjects ($\Delta V_{\text{within}}$) and EOG distance across subjects ($\Delta V_{\text{across}}$). Each of the four odorants was tested separately [iso-valeric acid (50%), l-carvone (50%), valeric acid (50%), linalool (50%), Table 1].

To further illustrate the clustering of responses within subjects, principal component analysis (PCA) was applied to the EOGs evoked by each odorant separately. The time window of a typical evoked EOG ($t_0 = 0.5\;s$ and $t_r = 2.5\;s$) consisted of 125 observations. To represent these in two dimensions, we applied the “princomp” function from the MatLab Statistics Toolbox (The data sets were Z scored before PCA was performed). The projections of the recordings on the first and second principal axes are plotted in Fig. 2.

Calculating the average area under the curve (AUC) of EOGs evoked by different odorant concentrations

To test for an EOG dose-response within subjects, we calculated the differences between the mean EOGs evoked by each of the odorants...
at the various concentrations (12.5, 25, 37.5, and 50%) and the blank response (0%) within subjects (Eq. 5). Prior to this calculation, we normalized the mean EOGs within subjects across concentrations to eliminate the effect of amplitude variation between subjects on the dose response effect measured within subjects (Eq 4, Fig. 3). We then performed a repeated-measures ANOVA followed by planned one-tailed tests with Benjamini-Hochberg correction for multiple comparisons. These differences are summarized in Table 4. The integration time window for the AUC calculation was ±15 time steps (480 ms) around the response latency, which was the time regime that showed best signal-to-noise ratio across subjects.

Subject’s average EOGs were normalized across concentrations (0, 12.5, 25, 37.5, and 50%) for each odorant separately before calculating their AUCs. For example, if an odorant elicited the amplitudes 0.01 mV, 0.05 mV, 0.09 mV, 0.11 mV, and 0.20 mV to the respective aforementioned concentrations in a subject, the normalized responses in that subject would now have the amplitudes: 0.05, 0.25, 0.45, 0.55, and 1, respectively. This normalization eliminated the effect of amplitude variation across subjects, and corrected the baseline to zero.

**Equation 4:** normalization of EOG curves across odorant concentrations

\[ V'_{Ci}(t) = \frac{V_{Ci}(t) - B_{i,C}}{B_{i,C} - N_{i,t}} \]

\( V'_{Ci}(t) \) is the normalized EOG recorded in subject i in response to an odorant at concentration C and at time t. \( V_{Ci}(t) \) is the mean EOG recorded in subject i to an odorant concentration C at time t. \( B_{i,C} \) is the baseline of the raw response, \( V_{Ci}(t) \) (4.8 s < t < 6 s). \( N_{i,t} \) is the maximal amplitude across concentrations recorded in subject i.

**Equation 5:** averaged area under an EOG curve

\[ \Delta V_{C_i-C_0} = \frac{1}{\Delta t \cdot N} \sum_{i=1}^{N} \sum_{t=0}^{t_{tf}} \left[ V_{C_i}(t) - V_{C_0}(t) \right] \]

\( \Delta V_{C_i-C_0} \) is the average difference between EOGs evoked by an odorant at concentration \( C_i \) and their blank, \( C_0 \) within subjects. \( \Delta V_{C_i-C_0} \) is also referred to as the average area under the curve (AUC) of the responses to \( C_i \). N is the number of subjects. \( V_{C_0}(t) \) is the average EOG recorded in subject i at time t in response to an odorant at concentration \( C_0 \). EOGs were normalized within subjects prior to this analysis (Eq 4). t is the integration index around the response latency, \( t_{tf} \). Notice that we took the latency of the evoked response to \( C_i \) and not to the blank that was often not distinguishable from noise. \( \Delta t \) normalizes this measure by the length of the integration window (30 time steps, 480 ms). An ANOVA followed by multiple comparison analysis was then performed between the AUCs of the different concentrations, namely, among \( \Delta V_{12.5-C_0} \), \( \Delta V_{25-C_0} \), \( \Delta V_{37.5-C_0} \) and \( \Delta V_{50-C_0} \) (Table 4).

**Stimulus classification of EOGs within subjects**

This manuscript focused on discriminating odorant concentrations in the EOG. However, we also conducted an exploratory analysis of our ability to discriminate odorant identity in the EOG. To test this, we took only EOGs evoked by the highest concentration of each odorant [i.e., either A iso-valeric acid (50%) vs. B l-carvone (50%) or A valeric acid (50%) vs. B linalool (50%)]. For each subject, each of the EOG responses was first assigned to its stimulating odorant, thus generating two odorant groups, A and B. We then tested all EOGs for their group of origin within subjects (leave 1 out process). To this end, we calculated the average least squared distance between the tested EOG curve and the remaining curves in each of the two groups in the time window of activation (\( t_f = 0.5 \text{s and } t_f = 2.5 \text{s}; \text{ Eq. 6} \)). This resulted in two scores of an estimated distance noted as \( D_{i,x}^{A,B} \) where i refers to the subject number, \( x \) refers to the tested EOG curve and A and B refer to the hypothesized group. The tested EOG curve was then assigned to the group it was closest to, namely, \( \min\{D_{i,x}^{A,B}\} \).

**Equation 6:** EOG classification within subjects

\[ D_{i,x}^{A,B} = \frac{1}{\Delta t \cdot n_0} \sum_{j=1}^{n_0} \sum_{t=0.5s}^{t=2.5s} \left| V_{i}(t) - V_{j}(t) \right| \]

A: \( V_{i}^{n_0} \)

B: \( V_{i}^{n_0} \)

Cost function for EOG classification within subjects: \( D_{i,x}^{A,B} \) is the average distance of the jth EOG curve recorded in subject i from the remaining EOGs in that subject evoked by odorant A. \( V_{i}(t) \) is the jth EOG curve recorded in subject i to odorant A with a total of \( n_0 \) such responses. \( V_{i}(t) \) is the tested EOG curve, \( x \), in the same subject. We wish to know whether x was evoked by odorant A or B. If \( D_{i,A} < D_{i,B} \), we decided that x was evoked by odorant A and visa versa. A and B specify the EOG curves attributed to each of the odorant groups, with a total of \( n_0 \) EOGs in group A and \( n_0 \) EOGs in group B. To calculate \( D_{i,B} \), one simply replaces the group index.
Stimulus classification from EOGs across subjects

Because of the increased variability in EOGs across subjects, stimulus classification across subjects was performed using normalized EOGs. Similar to the classification process within subjects, here too, we used only EOGs evoked by the highest odorant concentrations (50% olfactometer dilution) and attributed them to their stimulating odorant, which was one of the two response groups, A or B. The aspects that varied most between subjects were EOG amplitudes and their baselines. Therefore before odorant classification across subjects, all EOGs were normalized by their respective amplitudes and their baselines. Therefore before odorant classification across subjects, all EOGs were normalized by their respective amplitudes and their baselines. Therefore before odorant classification across subjects, all EOGs were normalized by their respective amplitudes and their baselines.

After normalization and grouping of the responses, we performed a “leave n out” process to test whether we can predict the odorant type from the remaining EOGs defined as the “learning set.” This normalization process did not depend on any external attribute such as the subject’s averaged responses or the odorant type. It only depended on the internal characteristics of the specific EOG curve that underwent normalization (i.e., its baseline and amplitude) and therefore could not bias the classification.

In the classification process, we iteratively discarded n EOGs in random order (defined as the “test set”) and tested whether we can predict their evoking odorant from their similarity to the remaining EOGs in each of the two groups, A and B (the learning set). n, the size of the test set, ranged between 1 and 10. At n = 1, the entire dataset was tested once; at n > 1, we iteratively discarded n recordings in random fashion such that each recording was discarded at least once. This procedure was repeated several times until the percent accuracy converged independently of the random choice of discarded responses. The size of the learning set (x axis in the bar plots of Fig. 7) is an outcome of the remaining responses in the database after the test set responses were discarded. Classification was assessed using the probability distributions of the two groups in the learning set. An example of such distribution plots is shown in Fig. 7A. The algorithm calculated the number of occurrences in which the tested curve fell in areas of positive likelihood for one of the odorants and zero likelihood for the other in the time window of a typical evoked response (t₀ = 0.5 s and t₁ = 2.5 s). The classification outcome was the odorant with higher number of positive likelihood occurrences.

Differences in EOG shape within subjects

To test for the differences in the shapes of the EOGs, three curve characteristics were extracted for the responses of each odorant at 50% olfactometer dilution: 1) the response amplitudes (N₁), 2) the response latency (L₁), and 3) a linear fit tangent to the EOG curve at the decay phase starting at L₁. These features are illustrated in Fig. 8A, top. Here too, the amplitudes (N₁) were extracted from the subject averaged EOGs using a minimum function between t₀ = 0.5 s and t₁ = 2.5 s and verified manually. The latencies (L₁) were derived from the amplitudes’ positions. Before calculating the linear fits of the EOG decay, we normalized the responses by their amplitudes as in the

Equation 7: EOG normalization across subjects

\[ V_{\text{norm}}^{\text{th}} = \frac{V_{\text{raw}}^{\text{th}} - B_{\text{th}}}{|B_{\text{th}} - N1_{\text{th}}|} \]

\( V_{\text{raw}}^{\text{th}} \) is the normalized EOG of the xth response in subject i. \( V_{\text{raw}}^{\text{th}} \) is the raw xth EOG recorded in subject i. \( B_{\text{th}} \) and \( N1_{\text{th}} \) are the baseline and amplitude of the raw EOG, \( V_{\text{raw}}^{\text{th}} \), respectively.

After normalization and grouping of the responses, we performed a “leave n out” process to test whether we can predict the odorant type from the remaining EOGs defined as the “learning set.” This normalization process did not depend on any external attribute such as the subject’s averaged responses or the odorant type. It only depended on the internal characteristics of the specific EOG curve that underwent normalization (i.e., its baseline and amplitude) and therefore could not bias the classification.

In the classification process, we iteratively discarded n EOGs in random order (defined as the “test set”) and tested whether we can predict their evoking odorant from their similarity to the remaining EOGs in each of the two groups, A and B (the learning set). n, the size of the test set, ranged between 1 and 10. At n = 1, the entire dataset was tested once; at n > 1, we iteratively discarded n recordings in random fashion such that each recording was discarded at least once. This procedure was repeated several times until the percent accuracy converged independently of the random choice of discarded responses. The size of the learning set (x axis in the bar plots of Fig. 7) is an outcome of the remaining responses in the database after the test set responses were discarded. Classification was assessed using the probability distributions of the two groups in the learning set. An example of such distribution plots is shown in Fig. 7A. The algorithm calculated the number of occurrences in which the tested curve fell in areas of positive likelihood for one of the odorants and zero likelihood for the other in the time window of a typical evoked response (t₀ = 0.5 s and t₁ = 2.5 s). The classification outcome was the odorant with higher number of positive likelihood occurrences.

Differences in EOG shape within subjects

To test for the differences in the shapes of the EOGs, three curve characteristics were extracted for the responses of each odorant at 50% olfactometer dilution: 1) the response amplitudes (N₁), 2) the response latency (L₁), and 3) a linear fit tangent to the EOG curve at the decay phase starting at L₁. These features are illustrated in Fig. 8A, top. Here too, the amplitudes (N₁) were extracted from the subject averaged EOGs using a minimum function between t₀ = 0.5 s and t₁ = 2.5 s and verified manually. The latencies (L₁) were derived from the amplitudes’ positions. Before calculating the linear fits of the EOG decay, we normalized the responses by their amplitudes as in the

Equation 7: EOG normalization across subjects

\[ V_{\text{norm}}^{\text{th}} = \frac{V_{\text{raw}}^{\text{th}} - B_{\text{th}}}{|B_{\text{th}} - N1_{\text{th}}|} \]

\( V_{\text{raw}}^{\text{th}} \) is the normalized EOG of the xth response in subject i. \( V_{\text{raw}}^{\text{th}} \) is the raw xth EOG recorded in subject i. \( B_{\text{th}} \) and \( N1_{\text{th}} \) are the baseline and amplitude of the raw EOG, \( V_{\text{raw}}^{\text{th}} \), respectively.

After normalization and grouping of the responses, we performed a “leave n out” process to test whether we can predict the odorant type from the remaining EOGs defined as the “learning set.” This normalization process did not depend on any external attribute such as the subject’s averaged responses or the odorant type. It only depended on the internal characteristics of the specific EOG curve that underwent normalization (i.e., its baseline and amplitude) and therefore could not bias the classification.

In the classification process, we iteratively discarded n EOGs in random order (defined as the “test set”) and tested whether we can predict their evoking odorant from their similarity to the remaining EOGs in each of the two groups, A and B (the learning set). n, the size of the test set, ranged between 1 and 10. At n = 1, the entire dataset was tested once; at n > 1, we iteratively discarded n recordings in random fashion such that each recording was discarded at least once. This procedure was repeated several times until the percent accuracy converged independently of the random choice of discarded responses. The size of the learning set (x axis in the bar plots of Fig. 7) is an outcome of the remaining responses in the database after the test set responses were discarded. Classification was assessed using the probability distributions of the two groups in the learning set. An example of such distribution plots is shown in Fig. 7A. The algorithm calculated the number of occurrences in which the tested curve fell in areas of positive likelihood for one of the odorants and zero likelihood for the other in the time window of a typical evoked response (t₀ = 0.5 s and t₁ = 2.5 s). The classification outcome was the odorant with higher number of positive likelihood occurrences.

Differences in EOG shape within subjects

To test for the differences in the shapes of the EOGs, three curve characteristics were extracted for the responses of each odorant at 50% olfactometer dilution: 1) the response amplitudes (N₁), 2) the response latency (L₁), and 3) a linear fit tangent to the EOG curve at the decay phase starting at L₁. These features are illustrated in Fig. 8A, top. Here too, the amplitudes (N₁) were extracted from the subject averaged EOGs using a minimum function between t₀ = 0.5 s and t₁ = 2.5 s and verified manually. The latencies (L₁) were derived from the amplitudes’ positions. Before calculating the linear fits of the EOG decay, we normalized the responses by their amplitudes as in the

Equation 7: EOG normalization across subjects

\[ V_{\text{norm}}^{\text{th}} = \frac{V_{\text{raw}}^{\text{th}} - B_{\text{th}}}{|B_{\text{th}} - N1_{\text{th}}|} \]

\( V_{\text{raw}}^{\text{th}} \) is the normalized EOG of the xth response in subject i. \( V_{\text{raw}}^{\text{th}} \) is the raw xth EOG recorded in subject i. \( B_{\text{th}} \) and \( N1_{\text{th}} \) are the baseline and amplitude of the raw EOG, \( V_{\text{raw}}^{\text{th}} \), respectively.

After normalization and grouping of the responses, we performed a “leave n out” process to test whether we can predict the odorant type from the remaining EOGs defined as the “learning set.” This normalization process did not depend on any external attribute such as the subject’s averaged responses or the odorant type. It only depended on the internal characteristics of the specific EOG curve that underwent normalization (i.e., its baseline and amplitude) and therefore could not bias the classification.

In the classification process, we iteratively discarded n EOGs in random order (defined as the “test set”) and tested whether we can predict their evoking odorant from their similarity to the remaining EOGs in each of the two groups, A and B (the learning set). n, the size of the test set, ranged between 1 and 10. At n = 1, the entire dataset was tested once; at n > 1, we iteratively discarded n recordings in random fashion such that each recording was discarded at least once. This procedure was repeated several times until the percent accuracy converged independently of the random choice of discarded responses. The size of the learning set (x axis in the bar plots of Fig. 7) is an outcome of the remaining responses in the database after the test set responses were discarded. Classification was assessed using the probability distributions of the two groups in the learning set. An example of such distribution plots is shown in Fig. 7A. The algorithm calculated the number of occurrences in which the tested curve fell in areas of positive likelihood for one of the odorants and zero likelihood for the other in the time window of a typical evoked response (t₀ = 0.5 s and t₁ = 2.5 s). The classification outcome was the odorant with higher number of positive likelihood occurrences. 
classification process across subjects (Eq. 7). This eliminated the effect of amplitude variation across subjects on the slope of the EOG decay. Linear curves were then fitted to the normalized responses in the time window between $L_1$ and $L_1 + \Delta t$, where $\Delta t$ was set to 40 time steps or 640 ms (Fig. 8A). This time window was about half the typical decay time (the time it takes the signal to reach baseline) and therefore ensured the goodness of fit.

RESULTS

Psychophysics

INCREASED CONCENTRATIONS WERE ASSOCIATED WITH INCREASED PERCEIVED INTENSITIES. Repeated-measures ANOVAs revealed that increasing odorant concentration resulted in increased perceived intensity across all four odorants [iso-valeric acid: $F(85,4) = 5.98, P < e^{-3}$; l-carvone: $F(85,4) = 15.94, P < e^{-9}$; valeric acid: $F(85,4) = 27.17, P < e^{-13}$ and linalool: $F(85,4) = 31.15, P < e^{-15}$]. The importance of this is in validating that consistent with previous results (Cain 1977a,b), increased olfactometer concentrations induced increase in the odorant’s perceived intensities (Fig. 4). Multiple comparisons between the normalized intensity estimates of the different concentrations is summarized in Table 3. Notably, the differences in perceived intensity were greater for linalool and valeric acid than for iso-valeric acid and l-carvone.

FIG. 6. Odorant classification from EOGs within subjects. Percent accuracy in EOG classification by the stimulating odorant at 100%> for either of the pairs. A: iso-valeric acid,50% olfactometer dilution) vs. l-carvone,50% olfactometer dilution) or B: valeric acid,50% olfactometer dilution) vs. linalool,50% olfactometer dilution). Classification was carried out by the minimal least squared distance between the tested EOG curve and the remaining EOGs within a subject’s dataset (Eq. 6).

FIG. 7. Odorant classification from EOGs across subjects. A: probability distributions of the evoked EOGs calculated from the normalized learning set (Eq. 7). Y axis is the evoked response in arbitrary units (au), X axis is the time course in seconds, stimulus duration = 500 ms starting at $t = 0.5$s. Top: valeric acid,50% olfactometer dilution) against linalool,50% olfactometer dilution) Bottom: iso-valeric acid,50% olfactometer dilution) against l-carvone,50% olfactometer dilution). Classification accuracy as a function of the size of the learning set. The tested odorants correspond to the probability plots of their learning sets shown in A.
Table 2. EOG amplitudes and latencies

<table>
<thead>
<tr>
<th></th>
<th>IVA</th>
<th>LC</th>
<th>T(18,2)</th>
<th>P</th>
<th>VA</th>
<th>Lin</th>
<th>T(22,2)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>N, mV</td>
<td>-0.149 ± 0.04</td>
<td>-0.223 ± 0.02</td>
<td>3.34</td>
<td>&lt;0.004</td>
<td>-0.185 ± 0.17</td>
<td>-0.265 ± 0.09</td>
<td>2.792</td>
<td>&lt;0.011</td>
</tr>
<tr>
<td>L, s</td>
<td>0.745 ± 0.04</td>
<td>0.852 ± 0.03</td>
<td>-1.45</td>
<td>&lt;0.17</td>
<td>0.8125 ± 0.04</td>
<td>0.7208 ± 0.03</td>
<td>0.764</td>
<td>&lt;0.46</td>
</tr>
</tbody>
</table>

Mean amplitudes (N, mV), latencies (L, s) and their respective SE. P for the test of differences in these attributes between the two odorants in each pair. These features were extracted from the raw responses in all subjects. Sample sizes: n(iso-valeric acid) = 41, n(L-carvone) = 43, n(valeric acid) = 51, n(linalool) = 58.

**EOG**

**EOG VARIABILITY WAS LOW WITHIN SUBJECTS BUT HIGH BETWEEN SUBJECTS.** The EOG distance between recordings obtained within subjects was significantly smaller than the EOG distance between recordings obtained from different subjects (all t > 3.7, all P < 0.002, Table 1). The averaged EOG distance within subjects was highest for L-carvone (0.021 mV, reflecting ~11% of the average response amplitude) and was lowest for valeric acid (0.036 mV, reflecting ~16% of the average response amplitude). The averaged EOG distance across subjects was highest for valeric acid (0.088 mV reflecting ~47% of the average response amplitude) and was lowest for L-carvone (0.055 mV, reflecting ~25% of the average response amplitude).

The projections of the raw EOG recordings on the first and second principal components are presented in the scatter plots of Fig. 2, colored by subject. For the odorant pair iso-valeric acid and L-carvone, the first principal component accounted for 59.9% of the explained variance, and the second principal component accounted for 14.5% of the explained variance. For the odorant pair valeric acid and L-carvone, the first principal component accounted for 85.1% of the explained variance.

Repeated-measures ANOVAs for all four odorants revealed a main effect of odorant concentration [iso-valeric acid: F(45,4) = 8.34, P < e-4; L-carvone: F(45,4) = 18.11, P < e-8; valeric acid: F(55,4) = 23.48, P < e-10 and linalool: F(55,4) = 34.38, P < e-10], reflecting an increase in the average area under the EOG curve (AUC) as a function of odorant concentration (Fig. 3). The only exception to this was a nonsignificant decrease in the AUC of L-carvone at 37.5% with respect to the mean AUC at 25% (Fig. 3B). Multiple comparison analysis between the AUCs of the responses evoked by different concentrations is summarized in Table 4. These tests suggested that the AUC of EOG scaled with odorant concentration. Furthermore, consistent with perceived intensity, linalool and valeric acid showed a more pronounced effect than iso-valeric acid and L-carvone as evident from the number of significant differences (Table 4).

The average AUCs of the odorants at various concentrations were correlated with their averaged perceived intensities [Fig. 5, r(Pearson) = 0.875, P < e-5].

**EOGs were concentration-dependent**

**FIG. 8. Different odorants elicited different EOG decay gradients.** A: typical example of subject normalized EOGs to two distinct odorants. au, arbitrary units. Top: L1—response latency (s); N1—response amplitude (equals –1 due to amplitude normalization). Continuous—iso-valeric acid, dashed—L-carvone. Bottom: continuous—valeric acid, dashed—linalool. B: averaged EOG decay slopes. Top: iso-valeric acid (continuous) and L-carvone (dashed), N = 10. Bottom: valeric acid (continuous) and linalool (dashed), N = 12. P is the result of a paired t-test between the decay slopes of the odorants.
and L-carvone were predicted correctly in size of 108 samples in the pair valeric acid/linalool. This result learning set size of $H_1^{11022}$ decreased with the decrease in learning set size (Fig. 7 $A$).

TABLE 4. Percent accuracy in EOG classification by the stimulating EOG responses, without prior normalization. Figure 6 shows subjects, stimulus classification was carried out using the raw within subjects Framework for odorant classification from EOGs $H_1^{11023}$.

In both odorant pairs, the percent accuracy of classification decreased with the decrease in learning set size (Fig. 7B). At a learning set size of $>100$ samples, the observed accuracy exceeded 90% with highest accuracy of 94.5% for learning set size of 108 samples in the pair valeric acid/linalool. This result implies that to distinguish between two odors by their evoked EOGs at $\geq 90\%$ accuracy, one needs a total of 50 repeats per odorant. Thus when recording five repeats from each stimulus, a total of 10 subjects is sufficient for distinction between EOGs of two different odors.

**Framework for odorant classification from EOGs within subjects**

Given the reduced variability in EOGs recorded in the same subjects, stimulus classification was carried out using the raw EOG responses, without prior normalization. Figure 6 shows the percent accuracy in EOG classification by the stimulating odorant at 50% olfactometer dilution within subjects. Isovaleric acid and valeric acid were predicted correctly in $>90\%$ of cases on average (binomial $P < 0.0001$), whereas linalool and L-carvone were predicted correctly in $>80\%$ of cases (binomial $P < 0.0005$).

**Framework for odorant classification from EOGs across subjects**

In the two odorant pairs, we classified EOGs by their evoking odorant at 88% accuracy or higher when the probability profile of the learning set was built on $\geq 84$ samples (Fig. 7B). Figure 7A illustrates the probability profiles of the learning set by which the tested EOGs were classified.

In both odorant pairs, the percent accuracy of classification decreased with the decrease in learning set size (Fig. 7B). At a learning set size of $>100$ samples, the observed accuracy exceeded 90% with highest accuracy of 94.5% for learning set size of 108 samples in the pair valeric acid/linalool. This result implies that to distinguish between two odors by their evoked EOGs at $\geq 90\%$ accuracy, one needs a total of 50 repeats per odorant. Thus when recording five repeats from each stimulus, a total of 10 subjects is sufficient for distinction between EOGs of two different odors.

**Different odors evoked different decay gradients**

When testing for differences in response amplitudes ($N1$), responses latencies ($L1$), and the gradients of the decay, we found significant differences in the amplitudes and in the decay slopes but not in the latencies of the EOGs between the odors of both odorant pairs (see Table 2 and Fig. 8). Specifically, the EOG responses to iso-valeric acid varied from the responses to L-carvone in amplitude [$N1(L_{carvone}) > N1(iso-valeric\ acid)$] in 9 of 10 subjects, binomial $P < 0.011$, but not in their latencies (Table 2). Similarly, the responses to valeric acid varied from the responses to linalool in amplitude [$N1(linalool) > N1(valeric\ acid)$] in 12 of 10 subjects, binomial $P < 0.02$, but not in latencies.

Decay gradients of the responses to iso-valeric acid varied from the gradients of L-carvone [$C1(iso-valeric\ acid) < C1(L_{carvone})$] in 8 of 10 subjects, binomial $P < 0.055$ or $F(73) = 10.596, P < 0.002$ and the decay gradients of the responses to Valeric Acid varied from the decay gradients of linalool [$C1(valeric\ acid) < C1(linalool)$] in 10 of 12 subjects, binomial $P < 0.02$.

**Multiple comparison analysis of the perceived intensities**

**TABLE 3.** Multiple comparison analysis of the perceived intensities

<table>
<thead>
<tr>
<th>Olfactometer Dilution, %</th>
<th>Linalool: $F(84,5) = 31.15$, $P &lt; e-15$</th>
<th>Valeric Acid: $F(84,5) = 27.17$, $P &lt; e-13$</th>
<th>Iso-Valeric Acid: $F(84,5) = 5.98$, $P &lt; e-3$</th>
<th>L-Carvone: $F(84,5) = 15.94$, $P &lt; e-9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>$P$</td>
<td>$h$</td>
<td>$t$</td>
</tr>
<tr>
<td>0</td>
<td>12.5</td>
<td>$&lt;e-04$</td>
<td>1</td>
<td>3.214</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>$&lt;e-09$</td>
<td>1</td>
<td>6.452</td>
</tr>
<tr>
<td>0</td>
<td>37.5</td>
<td>$&lt;e-13$</td>
<td>1</td>
<td>8.160</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>$&lt;e-14$</td>
<td>1</td>
<td>8.820</td>
</tr>
<tr>
<td>12.5</td>
<td>25</td>
<td>$&lt;e-005$</td>
<td>1</td>
<td>3.238</td>
</tr>
<tr>
<td>12.5</td>
<td>37.5</td>
<td>$&lt;e-05$</td>
<td>1</td>
<td>4.946</td>
</tr>
<tr>
<td>12.5</td>
<td>50</td>
<td>$&lt;e-06$</td>
<td>1</td>
<td>5.606</td>
</tr>
<tr>
<td>25</td>
<td>37.5</td>
<td>2.150</td>
<td>0.017</td>
<td>1.708</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>2.614</td>
<td>0.005</td>
<td>2.368</td>
</tr>
<tr>
<td>37.5</td>
<td>50</td>
<td>0.464</td>
<td>0.322</td>
<td>0</td>
</tr>
</tbody>
</table>

First two columns represent the olfactometer dilutions of the odorant concentrations being compared. Rejection criteria were corrected with Benjamini-Hochberg method to control for the false discovery rate. For each test the statistical contrast-$t$, its associated $P$ value and the test outcome are specified.
Discussion

EOGs were odorant-concentration dependent

That the EOG was odorant-concentration dependent was the key finding of this study (Fig. 3). When keeping the recording position fixed and the flow rate constant, EOG amplitude was concentration dependent in frogs and rats (Mozell 1962; Ottonson 1955; Scott et al. 2006). In the pioneering thesis describing human EOGs (Kobal 1981), similar concentration dependence was observed in a handful of test subjects, but this key feature has not been since systematically verified. Here we found that the area under the EOG curves reliably increased with odorant concentration (Table 4) in accordance with the odorant’s perceived intensity (Figs. 4 and 5). An additional encouraging marker as to the validity of this measure was that the odorant pairs with smaller differences in perceived intensity across concentrations were also the odorant pairs with smaller differences in EOG AUC (Tables 3 and 4, Fig. 5). That said, a weakness of this study was that perception and EOG were obtained from different subjects, and future efforts should bridge this gap. Previous EOG studies in the salamander and frog (Getchell 1974; Getchell and Shepherd 1978) suggested that this EOG AUC increase reflected an increase in the population size of responsive ORNs, and we hypothesize this to be the case here as well. That said, longer retention or slower clearance of the odorant from the mucosa, causing sustained activation of the same sized ORN population, remain plausible alternative explanations for this EOG AUC increase.

EOG variability was lower within than between subjects

As one might expect, we found that EOG variability within subjects was significantly lower than the EOG variability between subjects (Table 1, Eqs. 2 and 3). This was also illustrated in the clustering of the EOGs recorded in the same subjects shown in Fig. 2. This result may reflect the following methodological limitation: When recording from different subjects, one cannot identify the receptive tissue. Therefore recordings from different subjects can be theoretically treated as responses obtained from related though different receptive fields. To partially overcome this constraint in EOG feature extraction, we normalized the evoked responses by their respective amplitudes (Eq. 4). This eliminated the amplitude and baseline variations between subjects and captured the differences in the decay gradients of the responses.

EOGs and odorant classification

Although the focus of this manuscript was the representation of odorant concentration in the EOG, we also provided a test-of-concept analysis on EOG-based odorant classification. Several lines of evidence suggest that odor quality is reflected in differential patterning of spatial activation across the olfactory epithelium (OE). EOG studies from olfactory mucosa and from the olfactory nerve of insects, vertebrates, and mammals found odorant-specific patterns of activity across the OE (Edwards et al. 1988; Ezh et al. 1995; Getchell 1974; Getchell and Shepherd 1978; Kauer and Moulton 1974; Mackay-Sim and Kesteven 1994; Mozell 1962, 1964, 1966; Mustaparta 1971, 1990; Scott 2006; Scott et al. 1997, 2000, 1996; Thonmesen and Doving 1977). Similar results were obtained with voltage-sensitive dye studies in salamander and frog (Kendt and Mozell 1992; Kendt et al. 1995). These odorant-specific patterns may be attributed to either zonal receptor expression patterns (Buck 1996; Nef et al. 1992; Ressler et al. 1994; Strotmann et al. 1992; Sullivan and Dryer 1996; Vassar et al. 1993; Youngentob et al. 1995), odorant-solubility-related flow patterns (Mozell 1970; Mozell and Jagodowicz 1973; Mozell 1970), or a combination of both (Kimbell et al. 1997; Moulton 1976; Schoenfeld and Cleland 2005, 2006; Yang et al. 2007; Zhao et al. 2006).

Here we found that raw evoked EOGs can be classified by their evoking odorant within subjects (Fig. 6), and that following normalization by their amplitude, EOGs can also be classified by their evoking odorant across subjects (Fig. 7). We found that EOG amplitude as well as the gradient of decay varied between odorants, but their latencies remained the same (Table 2, Fig. 8B). Interestingly, the pleasant odorants (linalool and l-carvone) evoked larger amplitudes and faster decay gradients in comparison to their paired unpleasant odorants (valeric acid and iso-valeric acid, respectively). This may relate to the relative sorption and clearance rates of the odorants in the mucosa.

Our analysis on classification, however, can only be taken as an illustration for an approach on how to analyze such data. To make any conclusive arguments on odor classification from EOG, one must test more than two odorants within subjects. In turn, toward such an effort, a key result of this study is in defining the number of subjects and repeats typically needed to classify an EOG response between two odorants, namely ~10 subjects with approximately five successful recordings in each. To conclude, in spite of the inherent difficulties in the human EOG technique, this study provided an encoding platform for odor quantification and proposed such a platform for odorant classification, using evoked EOG recordings in awake behaving humans. To fully elucidate the mapping between the chemical and perceptual profile of odorants and their elicited spatial and temporal response on the OE with this recording technique, a large array of odorants should be tested at various known locations on the OE. Such an effort, combined with perceptual estimates obtained from human subjects, promises a unique window to olfactory coding.

Grants

This work was supported by the FP7 ideas grant 200850 from the European Research Council.

References


