



SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

SISSA Digital Library

A passive, camera-based head-tracking system for real-time, three-dimensional estimation of head position and orientation in rodents

*Original*

A passive, camera-based head-tracking system for real-time, three-dimensional estimation of head position and orientation in rodents / Vanzella, W.; Grion, N.; Bertolini, D.; Perissinotto, A.; Gigante, M.; Zoccolan, D.. - In: JOURNAL OF NEUROPHYSIOLOGY. - ISSN 0022-3077. - 122:6(2019), pp. 2220-2242. [10.1152/jn.00301.2019]

*Availability:*

This version is available at: 20.500.11767/110848 since: 2020-04-24T11:33:08Z

*Publisher:*

*Published*

DOI:10.1152/jn.00301.2019

*Terms of use:*

Testo definito dall'ateneo relativo alle clausole di concessione d'uso

*Publisher copyright*

note finali coverpage

(Article begins on next page)

1 **A passive, camera-based head-tracking system for real-time, 3D**  
2 **estimate of head position and orientation in rodents**

3  
4  
5 **Walter Vanzella<sup>1,2,\*</sup>, Natalia Grion<sup>1,\*</sup>, Daniele Bertolini<sup>1,a,\*</sup>, Andrea Perissinotto<sup>1,2,b</sup>**  
6 **and Davide Zoccolan<sup>1</sup>**

7  
8 *1. Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Trieste,*  
9 *34136, Italy*

10  
11 *2. Glance Vision Tech. Srl, Area Science Park, Ed. Q, 34149 Trieste, Italy*

12  
13 *\*These authors contributed equally to this work*

14  
15 *a. Current address: VIVISOL (SOL group), Milan, Italy*

16 *b. Current address: Department of Biomedical Engineering, Tel Aviv University, Ramat*  
17 *Aviv, Israel*

18  
19 **Corresponding author**

20  
21 **Davide Zoccolan**

22 International School for Advanced Studies (SISSA)

23 Via Bonomea, 265 34136 Trieste (TS) ITALY

24 Office: +39 040 3787 701 Fax: +39 040 3787 702

25 E-mail: [zoccolan@sissa.it](mailto:zoccolan@sissa.it)

## 32 **Abstract**

33 Tracking head's position and orientation of small mammals is crucial in many behavioral  
34 neurophysiology studies. Yet, full reconstruction of the head's pose in 3D is a  
35 challenging problem that typically requires implanting custom headsets made of multiple  
36 LEDs or inertial units. These assemblies need to be powered in order to operate, thus  
37 preventing wireless experiments, and, while suitable to study navigation in large arenas,  
38 their application is unpractical in the narrow operant boxes employed in perceptual  
39 studies. Here we propose an alternative approach, based on passively imaging a 3D-  
40 printed structure, painted with a pattern of black dots over a white background. We show  
41 that this method is highly precise and accurate and we demonstrate that, given its  
42 minimal weight and encumbrance, it can be used to study how rodents sample sensory  
43 stimuli during a perceptual discrimination task and how hippocampal place cells  
44 represent head position over extremely small spatial scales.

45

## 46 **Introduction**

47 Careful monitoring and quantification of motor behavior is essential to investigate a  
48 range of cognitive functions (such as motor control, active perception and spatial  
49 navigation) in a variety of different species. Examples include tracking eye movements in  
50 primate and non-primate species (Rommel, 1984; Stahl et al., 2000; Zoccolan et al., 2010;  
51 Kimmel et al., 2012; Wallace et al., 2013; Payne and Raymond, 2017), monitoring  
52 whisking activity in rodents (Knutsen et al., 2005; Perkon et al., 2011; Rigosa et al.,  
53 2017), and tracking the position of virtually any species displaying interesting navigation  
54 patterns – from bacteria (Berg and Brown, 1972) and invertebrate species (Mazzoni et al.,  
55 2005; Garcia-Perez et al., 2005; Mersch et al., 2013; Cavagna et al., 2017), to small  
56 terrestrial (Tort et al., 2006; Aragão et al., 2011) and aerial mammals (Tsoar et al., 2011;  
57 Yartsev and Ulanovsky, 2013) and birds (Attanasi et al., 2014). In particular, studies in  
58 laboratory animals aimed at measuring the neuronal correlates of a given behavior require  
59 tools that can accurately track it in time and space and record it along with the underlying  
60 neuronal signals.

61 A classical application of this approach is to track the position of a light-emitting  
62 diode (LED), mounted over the head of a rat or a mouse, while recording the activity of  
63 place cells in hippocampus (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978)  
64 or grid cells in entorhinal cortex (Fyhn et al., 2004; Hafting et al., 2005), with the goal of  
65 understanding how space is represented in these brain structures (Moser et al., 2008,  
66 2015). It is also common, in studies of spatial representations, to track the *yaw* of the  
67 head (i.e., its orientation in the horizontal plane where the rodent navigates; see Figure  
68 2C), to investigate the tuning of neurons in hippocampus, entorhinal cortex and other  
69 limbic structures for head direction, speed or angular velocity (Sargolini et al., 2006;  
70 Taube, 2007; Kropff et al., 2015; Acharya et al., 2016). In these applications, the yaw is  
71 tracked through an overhead video camera imaging two LEDs of different colors (e.g.,  
72 red and green), mounted over the head stage of the neuronal recording system, and placed  
73 along the anteroposterior axis of the head. In some studies, this LED arrangement was  
74 also used to estimate the *pitch* of the head (i.e., its rotation about the interaural axis; see  
75 Figure 2C), by measuring the distance between the two LEDs in the image plane  
76 (Stackman and Taube, 1998; Bassett and Taube, 2001).

77 It is more difficult (and only rarely it has been attempted) to achieve a complete  
78 estimate of the pose and location of the head in the three-dimensional (3D) space – i.e., to  
79 simultaneously track the three Cartesian coordinates of the head and the three Euler  
80 angles that define its orientation: yaw, pitch and *roll* (with the latter defined as the  
81 rotation about the head's anteroposterior axis; see Figure 2C). Recently, two groups have  
82 successfully tracked in 3D the head of small, freely-moving mammals through  
83 videography, by relying either on a single camera imaging a custom tetrahedral  
84 arrangement of four LEDs with different colors (Finkelstein et al., 2015), or on multiple  
85 cameras (up to four) imaging custom 3D arrangements of up to six infrared (IR) LEDs  
86 (Sawinski et al., 2009; Wallace et al., 2013). Other groups have used instead inertial  
87 measurement units (IMUs), such as accelerometers and gyroscopes, mounted over the  
88 head of a rat, to record its angular displacement and velocity along the three Euler  
89 rotation axes (Pasquet et al., 2016; Kurnikova et al., 2017).

90 All these approaches provide accurate measurements of head position and pose in  
91 3D. However, having been developed as ad hoc solutions for specific experimental

92 settings, their design is not necessarily optimal for every application domain. For  
93 instance, most of these systems were conceived to track the head of small mammals  
94 roaming over an open-field arena, where the relatively large size of the custom LED- or  
95 IMU-based headset (extending several cm above and/or around the animal's head) was  
96 not a issue in terms of encumbrance or obstruction. Moreover, these headsets need to be  
97 powered in order to operate. In general, this requires dedicated wires, which increase the  
98 stiffness of the bundle of cables connected to the headstage of the recording system, and  
99 prevent performing fully unplugged recordings using headstages equipped with wireless  
100 transmitters (Szuts et al., 2011; Pinnell et al., 2016).

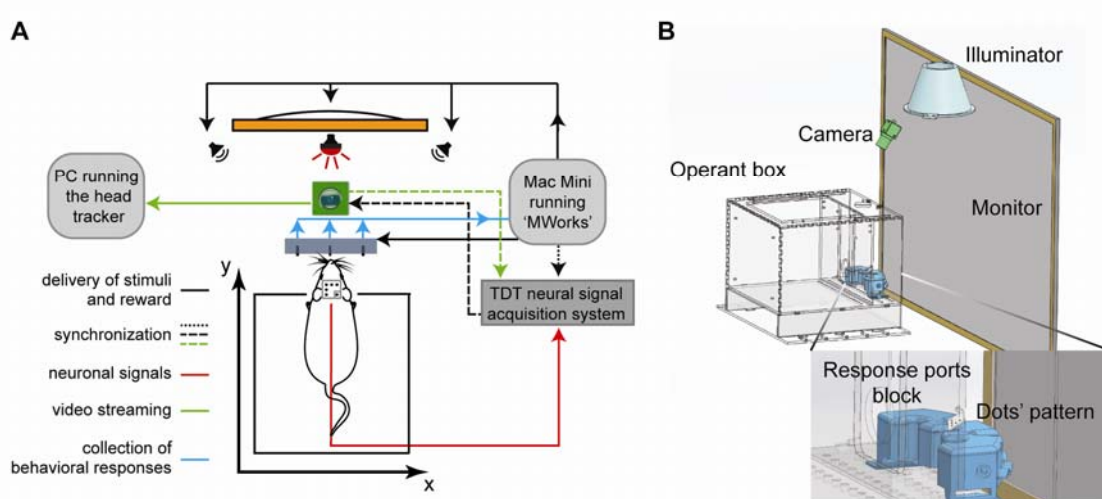
101 In this study, we tried to overcome these limitations, by using a single, overhead  
102 camera to passively image a 3D-printed structure, painted with a pattern of black dots  
103 over a white background and mounted over the head of a rat. The small size of the pattern  
104 (1.35x1.35x1.5 cm) makes it ideal for perceptual studies, where a rodent performs a  
105 discrimination task inside a narrow operant box, often with its head inserted through an  
106 opening or confined within a funnel, as in the studies of rodent visual perception recently  
107 carried out by our group (Zoccolan et al., 2009; Tafazoli et al., 2012; Alemi-Neissi et al.,  
108 2013; Rosselli et al., 2015; Nikbakht et al., 2018; Djurdjevic et al., 2018) and other  
109 authors (Vermaercke and Op de Beeck, 2012; Horner et al., 2013; Mar et al., 2013;  
110 Kurylo et al., 2015; De Keyser et al., 2015; Bossens et al., 2016; Stirman et al., 2016;  
111 Kurylo et al., 2017; Yu et al., 2018). In what follows, beside describing in details the  
112 equipment and the algorithm upon which our method is based (Materials and Methods)  
113 and validate its accuracy and precision (first part of the Results and discussion), we  
114 provide a practical demonstration of the way our head tracker can help understanding: 1)  
115 how a rat samples the sensory stimuli during a visual or auditory discrimination task; and  
116 2) how hippocampal neurons represent head position over extremely small spatial scales  
117 around the area where the animal delivers its perceptual decision and collects the reward.

118

## 119 **Results and discussion**

120 Our head tracker was developed in the context of a two-alternative forced-choice (2AFC)  
121 perceptual discrimination experiment (Zoccolan, 2015; Zoccolan and Di Filippo, 2018),

122 involving the visual and auditory modalities. A scheme of the experimental set up is  
123 shown in Figure 1A. The rats performed the discrimination task inside an operant box,  
124 similar to the one used in (Djurdjevic et al., 2018), and equipped with a monitor and two  
125 speakers for the presentation of the sensory stimuli. The task required each animal to  
126 insert the head through an opening in the wall facing the monitor and interact with an  
127 array of three response ports (i.e., three feeding needles, equipped with proximity  
128 sensors). Specifically, the rat had to lick the central needle to trigger the presentation of  
129 the stimulus. Afterward, he had to lick one of the lateral needles to report his perceptual  
130 choice and receive a liquid reward, in case of successful discrimination (see Materials  
131 and Methods).



132

133 **Figure 1.** Illustration of the experimental rig where perceptual discrimination tests were  
134 combined with in-vivo neuronal recordings and real-time tracking of the head of a rat  
135 subject. (A) The animal was placed in an operant box and learned to protrude his head  
136 through an opening in one of the walls, so as to face the stimulus display (orange  
137 rectangle) and an array of response ports (gray rectangle), which also delivered liquid  
138 reward. Stimulus presentation and reward delivery (solid black arrows), as well as  
139 collection of the behavioral responses (solid cyan arrow) were controlled by the freeware  
140 application MWorks, running on a Mac mini (light-gray, rounded box on the right).  
141 MWorks also streamed unique codes (dotted black arrow) with the identity of the  
142 presented stimuli to the PC that controlled the recording of the neurophysiological  
143 signals from the rat brain (red arrow) through a Tucker-Davis Technologies (TDT)  
144 amplification/acquisition system (dark-gray box on the right). The head-tracking software

145 run on a dedicated PC (light-gray, rounded box on the left), which received the video  
146 stream (solid green arrow) outputted by an industrial monochromatic CMOS camera  
147 (green box) placed above the operant box, along with a far-red light source (red bulb).  
148 The camera imaged a 3D pattern of dots mounted over the head of the animal (see  
149 Figure 2 for details). Image uptake was triggered by the TDT system (dashed black  
150 arrow), which, in turn, received from the camera a unique identification code for every  
151 acquired frame (dashed green arrow). The same code was also saved by the head-  
152 tracking software. The x and y axes (thick black arrows) indicate the Cartesian plane  
153 corresponding to the floor of the operant box. **(B)** A CAD rendering of some of the key  
154 elements of the experimental rig. The drawing allows appreciating the relative size and  
155 position of the operant box (light gray), stimulus display (dark gray), camera (dark green)  
156 and illuminator (light green). The inset shows a detail of the 3D-printed block (cyan)  
157 holding the response ports and allows appreciating the size and typical position of the  
158 dot's pattern, relative to the other components of the rig.

159

160 Presentation of the visual and auditory stimuli, collection of the behavioral  
161 responses, delivery of reward, and real-time control of the flow of events during the  
162 experiment was achieved with the freeware, open source application MWorks (Figure  
163 1A, solid black and cyan arrows), but any other software that allows programing  
164 psychophysical experiments would be suitable. Stimulus presentation was synchronized  
165 with the amplification/acquisition system for electrophysiological recordings, a Tucker-  
166 Davis Technologies (TDT) recording system (dotted black arrow) – i.e., at every trial,  
167 MWorks sent a code with the identity of the stimulus via UDP to the acquisition system.

168 The head tracker consisted of a far-red light source (red bulb) and an  
169 industrial monochromatic CMOS camera (green box), both placed above the operant box,  
170 with the camera feeding its video stream (green solid arrow) to a dedicated PC, equipped  
171 with the head tracking software. To synchronize the head tracker with the TDT  
172 acquisition system, we adopted a master-slave configuration, where the latter worked as a  
173 master, generating a square wave that triggered the camera image uptake (black dashed  
174 line). The camera, in turn, generated a unique identification code for every acquired  
175 frame, which was both saved by the PC running the head-tracking software (along with

176 the Cartesian coordinates and pose of the head) and fed back to the TDT (green dashed  
177 arrow). In this way, the acquisition system simultaneously recorded the electric activity  
178 from the brain (red line), the identity and time of presentation of the stimuli displayed by  
179 MWorks, and the time and frame number of each image acquired by the overhead  
180 camera. Figure 1B shows a CAD drawing with a scaled representation of the key  
181 components of the rig and their relative position: the operant box (light gray), the monitor  
182 (dark gray), the block holding the feeding needles (cyan), the overhead camera (dark  
183 green) and the case of the light source (light green).

184 The key element of our head tracking systems is a 3D pattern of dots, mounted  
185 over the head of the animal and imaged by the overhead camera. The pattern is a small  
186 ((1.35x1.35x1.5 cm), light-weight (1.10 grams approximately), easy-to-place 3D printed  
187 structure, with 5 coplanar dots on a white background located over a plate, plus a sixth,  
188 elevated dot placed over a pillar (Figure 2A). At the beginning of each experimental  
189 session, the pattern is mounted on top of the head of the animal (Figure 2B), using an  
190 apposite magnet that connects it to a base that was previously surgically implanted  
191 (Materials and Methods). As described in detail in Materials and Methods, the head-  
192 tracking algorithm detects the 6 black dots in the frames of the video stream and  
193 computes the position and pose of the pattern in the 3D space of the camera reference  
194 system. More specifically, a *pattern reference system* ( $x'$ ,  $y'$ ,  $z'$ ; purple arrows in Figure  
195 2C) is defined, with the origin  $O'$  placed over one of the dots, the  $x'$  and  $y'$ s axes parallel  
196 to the edges of the plate, and the  $z'$  axis perpendicular to it (i.e., parallel to the pillar). In  
197 the ideal case in which the pattern had been precisely aligned to the anatomical axes of  
198 the head at the time of the implant,  $x'$  corresponds to the head's anteroposterior axis,  
199 while  $y'$  corresponds to the head's interaural axis. The *camera reference system* ( $x$ ,  $y$ ,  $z$ ;  
200 black arrows in Figure 2C) results instead from calibrating the camera using a standard  
201 procedure that consists in imaging a checkerboard pattern placed at various positions and  
202 orientations (see Materials and Methods). Once the camera is calibrated, the head-  
203 tracking algorithm provides: 1) the three Cartesian coordinates of  $O'$  in the camera  
204 reference system; and 2) the rotation matrix  $R$  that defines the 3D rotation bringing the  
205 pattern reference system ( $x'$ ,  $y'$ ,  $z'$ ) to be aligned with the camera reference system ( $x$ ,  $y$ ,  
206  $z$ ).  $R$  is defined as:



207

$$208 \quad R = R_z(\varphi_c) R_y(\theta_c) R_x(\gamma_c), \quad (1)$$

209

210 where  $R_z$ ,  $R_y$  and  $R_x$  are the elemental rotation matrixes that define intrinsic rotations  
211 by the Euler angles  $\varphi_c$  (yaw),  $\theta_c$  (pitch), and  $\gamma_c$  (roll) about the axes of the camera  
212 reference system. More specifically:

213

$$R_z(\varphi_c) = \begin{bmatrix} \cos \varphi_c & -\sin \varphi_c & 0 \\ \sin \varphi_c & \cos \varphi_c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

214

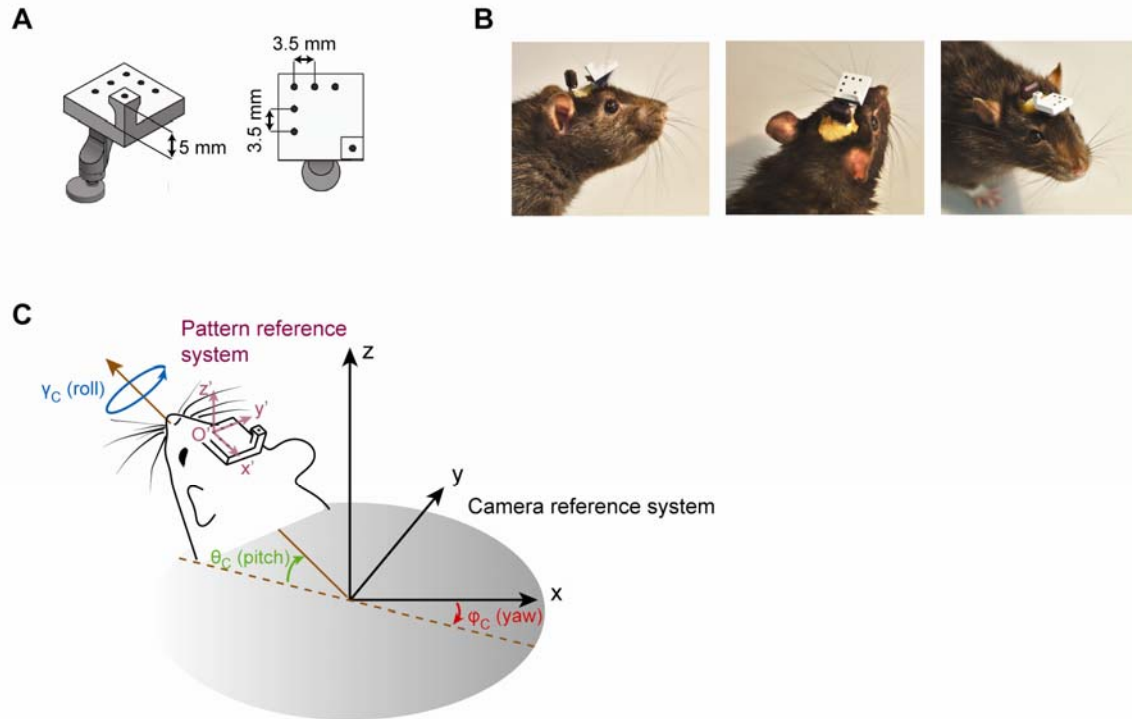
$$R_y(\theta_c) = \begin{bmatrix} \cos \theta_c & 0 & \sin \theta_c \\ 0 & 1 & 0 \\ -\sin \theta_c & 0 & \cos \theta_c \end{bmatrix}$$

215

$$R_x(\gamma_c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma_c & -\sin \gamma_c \\ 0 & \sin \gamma_c & \cos \gamma_c \end{bmatrix},$$

216

217 where, with reference to Figure 2C:  $\varphi_c$  is the angle between the projection of  $x'$  onto the  
218 camera  $(x, y)$  plane and the camera  $x$  axis;  $\theta_c$  is the angle between  $x'$  and the  $(x, y)$  plane;  
219 and  $\gamma_c$  is the rotation angle of the pattern around  $x'$ .



220

221 **Figure 2.** Illustration of the 3D pattern of dots used for head tracking and of the Euler  
222 angles that define its pose in the camera reference system. (A) A CAD rendering of the  
223 pattern of dots that, once mounted over the head of a rat, allows its position and pose to  
224 be tracked. Notice that, among the 6 dots, five are coplanar, while the sixth is placed  
225 over an elevated pillar. Also notice the arm that allows the pattern to be attached to a  
226 matching base surgically implanted over the skull of the rat. (B) Different views of a rat  
227 with the dots' pattern mounted over his head. (C) Definition of the angles of rotation of  
228 the reference system centered on the dots' pattern ( $x'$ ,  $y'$ ,  $z'$ ; purple arrows) with respect  
229 to the camera reference system ( $x$ ,  $y$ ,  $z$ ; black arrows). The three Euler angles – yaw,  
230 pitch and roll – are shown, respectively, by the red, green and blue arrows.  $O'$  indicates  
231 the origin of the pattern reference system. The brown arrow indicates the direction where  
232 the rat's nose is pointing and it is parallel to the head's anteroposterior axis (i.e., to the  $x'$   
233 axis). The dashed brown line is the projection of the brown arrow over the ( $x$ ,  $y$ ) plane of  
234 the camera reference system.

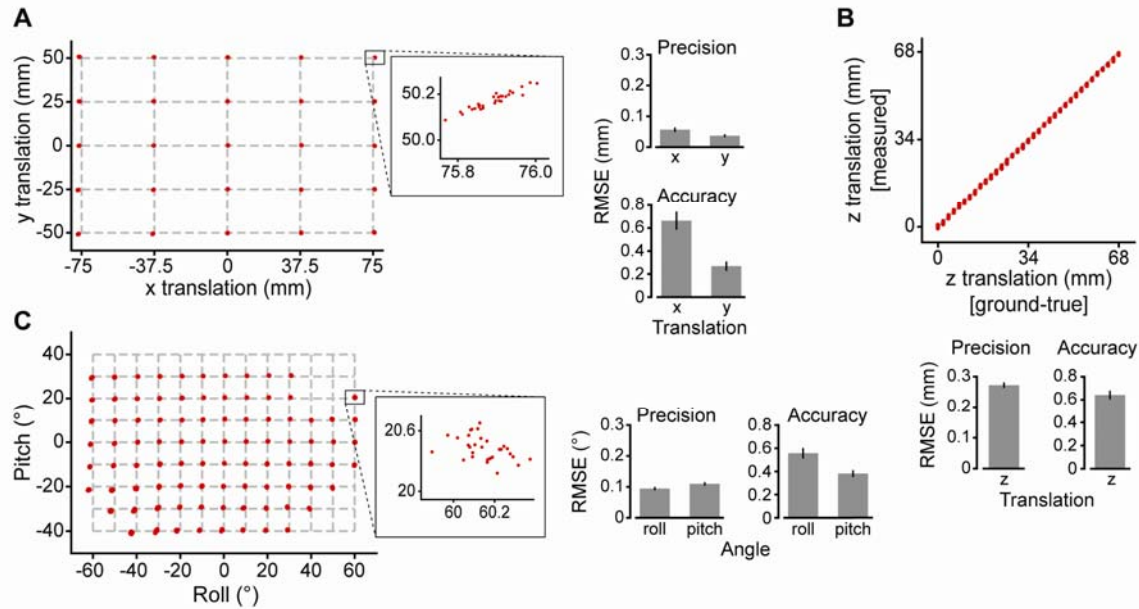
235

236 It should be noticed that the three Euler angles  $\varphi_C$ ,  $\theta_C$ , and  $\gamma_C$ , as well as the three  
237 Cartesian coordinates of  $O'$ , are not immediately applicable to know the pose and  
238 position of the head in the environment. First, they are relative to the camera reference

239 system, while the experimenter needs to know them with respect to some meaningful  
240 environmental landmarks (e.g., the floor and walls of the arena or the operant box, where  
241 the animal is tested). Second, no matter how carefully the pattern is placed over the head  
242 of the rat, in general, the  $(x', y', z')$  axes will not be perfectly aligned to the anatomical  
243 axes of the head – e.g.,  $x'$  and  $y'$  will not be perfectly aligned to the head's  
244 anteroposterior and interaural axes. However, expressing the pose and position of the  
245 pattern in camera coordinates allows measuring the nominal precision and accuracy that  
246 the head tracker can achieve, which is essential to validate the system. In the next section,  
247 we illustrate how we collected these validation measurements, while, in the following  
248 section, we explain how the actual position and pose of the head can be expressed in a  
249 convenient reference system, by collecting images of the checkerboard and dot patterns  
250 at, respectively, a reference position and pose.

### 251 **Validation of the head tracker: nominal precision and accuracy in the camera** 252 **reference system**

253 To measure the precision and accuracy of the head tracker (Figure 3), we used a custom  
254 combination of breadboards, linear stages, rotary stages and goniometers to hold the dots'  
255 pattern in known 3D positions and poses. This allowed comparing the ground-true  
256 coordinates/angles of the pattern with the measurements returned by the head tracker. For  
257 these measurements, the camera was positioned in such a way to have its optical axis  
258 perpendicular to the floor of the testing area.



259

260 **Figure 3.** Nominal precision of the head-tracker in the camera reference system. (A)  
261 Left: validation measurements (red dots) obtained by placing the dots' pattern on a grid  
262 of 5 x 5 ground-truth positions (grid's intersections), over the floor of the testing area,  
263 using the breadboard shown in Figure 3-figure supplement 1A. Although 30 repeated  
264 head-tracker measurements were taken at each tested position, the spread of the  
265 measurements around the grid intersections is not appreciable, because of the very high  
266 precision of the head tracker. Only by zooming into the area of the grid intersection at  
267 the sub-millimetre scale, the spread of the red dots becomes visible (inset). Right: mean  
268 precision (top) and accuracy (bottom) of the head-tracker measurements over the 25  
269 tested positions, as estimated by computing the RMSE relative either to the mean of  
270 each set of 30 repeated measurements (precision) or to the ground-truth positions  
271 (accuracy). Note that the sets of measured and ground-truth positions were aligned  
272 using Procrustes analysis (see main text). (B) Top: validation measurements (red dots)  
273 obtained by vertically displacing the dots' pattern (relative to the floor of the testing area)  
274 of 34 consecutive increments, using the stereotax arm shown in Figure 3-figure  
275 supplement 1B. Again, given their high precision, the spread of the 30 repeated  
276 measurements taken at each ground-truth value is barely appreciable. Bottom: mean  
277 precision (left) and accuracy (right) of the head-tracker measurements over the 36 tested  
278 vertical displacements (RMSE computed as in A). (C) Left: validation measurements  
279 (red dots) obtained by setting the roll and pitch angles of the dots' pattern to a  
280 combination of 13 x 9 ground-truth values (grid's intersections), using the custom

281 assembly shown in Figure 3-figure supplement 1C. The inset allows appreciating the  
282 spread of the 30 measurements taken at one of the tested angle combinations. Note  
283 that, for some extreme rotations of the pattern, no measurements could be taken (grid's  
284 intersections without red dots), since the dots on the pattern were not visible. Right:  
285 mean precision (left) and accuracy (right) of the head-tracker measurements over the set  
286 of tested angle combinations (RMSE computed as in **A**). Note that the sets of measured  
287 and ground-truth angle combinations were optimally aligned following the procedure  
288 described in (Wallace et al., 2013) (see main text).

289

290 We first measured the ability of the system to track two-dimensional (2D)  
291 displacements of the pattern over the floor of the testing area. To this aim, the pattern was  
292 placed over a 3D printed breadboard with a grid of 5 x 5 holes (37.5 mm and 25 mm  
293 apart along, respectively, the horizontal and vertical dimensions; Figure 3-figure  
294 supplement 1A). For each of the 25 breadboard locations, we took a set of 30 repeated,  
295 head-tracker measurements of the origin of the pattern in the camera reference system  
296 (i.e.,  $O'$  in Figure 2C). Since the coordinates of the grid holes in such a reference system  
297 are not known a priori, a Procrustes analysis (Gower et al., 2004) was applied to find the  
298 optimal match between the set of 25x30 measures returned by the head tracker, and the  
299 known, physical positions of the holes of the breadboard. Briefly, the Procrustes analysis  
300 is a standard procedure to optimally align two shapes (or two sets of points, as in our  
301 application) by uniformly rotating, translating and scaling one shape (or one set of points)  
302 with respect to the other. In our analysis, since we compared physical measurements  
303 acquired with a calibrated camera, we did not apply the scale transformation (i.e., the  
304 scale factor fixed to 1). When applied to our set of 25x30 measures, the Procrustes  
305 analysis returned a very good match with the set of 25 ground-true positions of the grid  
306 (Figure 3A, left; red dots vs. grid intersections). As shown by the virtually absent spread  
307 of the dots at each intersection, the root mean square error (RMSE) of each set of 30  
308 measurements, relative to their mean, was very low, yielding a mean *precision* (across  
309 positions) of  $0.056 \pm 0.007$  mm along the x axis and  $0.037 \pm 0.005$  mm along the y axis  
310 (Figure 3A, right; top bar plot). The RMSE of each set of 30 measurements, relative to  
311 the corresponding grid intersection, was also very low, yielding a mean *accuracy* (across

312 positions) of  $0.663 \pm 0.079$  mm along the x axis and  $0.268 \pm 0.041$  mm along the y axis  
313 (Figure 3A, right; bottom bar plot). To verify the ability of the head-tracker to estimate  
314 the height of the pattern (i.e., its displacement along the z axis) when it was varied over a  
315 range of positions, the pattern was mounted on a stereotaxic arm through a 3D-printed  
316 custom joint (Figure 3-figure supplement 1B). The arm was positioned in such a way to  
317 be perpendicular to the floor of the testing area, thus allowing the pattern to be displaced  
318 vertically of 34 consecutive 2 mm increments. After each increment, the z coordinate of  
319 the pattern was estimated by the head tracker in 30 repeated measurements, which were  
320 then compared to the total physical displacement of the stereotaxic arm up to that point  
321 (Figure 3B, top). The resulting estimates were again very precise and accurate (Figure  
322 3B, bottom), with RSME values that were close to those obtained previously for the x  
323 and y displacements (compare the z bars of Figure 3B to the x and y bars of Figure 3A).

324 For the validation of the angular measurements, we built a custom assembly,  
325 made of a rotary stage mounted on a stereotaxic arm, that allowed rotating the pattern  
326 about two orthogonal axes, corresponding to roll and pitch (Figure 3-figure supplement  
327 1C). Rotations about each axis were made in  $10^\circ$  steps, spanning from  $-60^\circ$  to  $60^\circ$  roll  
328 angles and from  $-40^\circ$  to  $40^\circ$  pitch angles, while the yaw was kept fix at  $0^\circ$ . Again, 30  
329 repeated head-tracker measurements were collected at each know combination of angles  
330 over the resulting 13 x 9 grid. As for the case of the 2D displacements, the angles  
331 returned by the head tracker were not immediately comparable with the nominal rotations  
332 on the rotary stages, because the two sets of angles are measured with respect to two  
333 different reference systems – i.e., the camera reference system (as defined in Figure 2C),  
334 and the stage reference system (as defined by the orientation of the pattern in the physical  
335 environment, when the rotations of the stages are set to zero). Therefore, in order to  
336 compare the nominal rotations of the pattern on the stages ( $R_{\text{stages}}^{\text{nom}}$ ) to their estimates  
337 provided by the head-tracker ( $R_{\text{cam}}^{\text{est}}$ ), we first had to express the former in the camera  
338 reference system ( $R_{\text{cam}}^{\text{nom}}$ ). To this aim, we followed the same approach of (Wallace et al.,  
339 2013) and we computed  $R_{\text{cam}}^{\text{nom}}$  as:

340

$$341 \quad R_{\text{cam}}^{\text{nom}} = R_{\text{stage}}^{\text{cam}} \cdot R_{\text{stages}}^{\text{nom}} \cdot (R_{\text{stage}}^{\text{cam}})^T \cdot R_{\text{cam}}^{\text{nom}0}, \quad (2)$$

342

343 where  $R_{\text{cam}}^{\text{nom}0}$  is the pose of the pattern (in the camera reference system) when all the  
344 rotations of the stages are nominally set to zero (i.e., reference rotation), and  $R_{\text{stage}}^{\text{cam}}$  is the  
345 matrix mapping the stage reference system into the camera reference system [note that  
346 each matrix in (2) is in the form shown in (1)]. The matrixes  $R_{\text{stage}}^{\text{cam}}$  and  $R_{\text{cam}}^{\text{nom}0}$  are  
347 unknown that can be estimated by finding the optimal match between the nominal and  
348 estimated rotations in camera coordinates, i.e., between  $R_{\text{cam}}^{\text{nom}}$ , as defined in (2), and  $R_{\text{cam}}^{\text{est}}$ .  
349 Following (Wallace et al., 2013), we defined the rotation difference matrix  $R_{\text{diff}} = R_{\text{cam}}^{\text{est}} \cdot$   
350  $R_{\text{cam}}^{\text{nom}T}$ , from which we computed the error of a head-tracker angle estimate as the total  
351 rotation in  $R_{\text{diff}}$ , i.e., as:

352

$$err = \cos^{-1} \left( \frac{\text{trace}(R_{\text{diff}}) - 1}{2} \right).$$

353

354 By minimizing the sum of  $err^2$  over all tested rotations of the stages (using Matlab  
355 *fminsearch* function), we obtained a very close match between estimated and nominal  
356 stage rotations. This is illustrated in Figure 3C (left), where the red dots are the head-  
357 tracker estimates and the grid intersections are the nominal rotations. As for the case of  
358 the Cartesian displacements, also for the pitch and roll angles, the head tracker returned  
359 very precise (roll:  $\text{RSME} = 0.095 \pm 0.005^\circ$ ; pitch:  $\text{RSME} = 0.109 \pm 0.006^\circ$ ) and accurate  
360 (roll:  $\text{RSME} = 0.557 \pm 0.045^\circ$ ; pitch:  $\text{RSME} = 0.381 \pm 0.030^\circ$ ) measurements (Figure 3C,  
361 right).

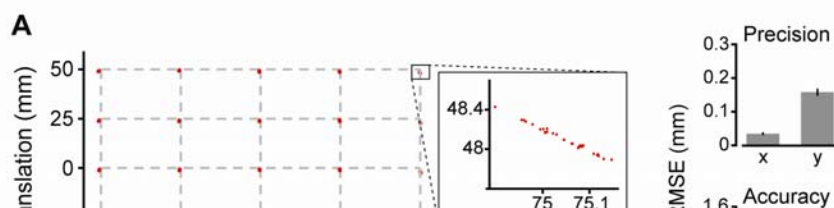
### 362 **Operation of the head tracker: measuring displacements and rotations relative to** 363 **reference poses in the physical environment**

364 While the validation procedure described in the previous section provides an estimate of  
365 the nominal precision and accuracy of the head tracker, measuring Cartesian coordinates  
366 and Euler angles in the camera reference system is impractical. To refer the head tracker  
367 measurements to a more convenient reference system in the physical environment, we 3D  
368 printed a custom adapter to precisely place the checkerboard pattern used for camera  
369 calibration over the block holding the feeding needles, in such a way to be parallel to the  
370 floor of the operant box, with the vertex of the top-right black square vertically aligned

371 with the central needle (see Figure 4-figure supplement 1A). We then acquired an image  
372 of this *reference checkerboard*, which served to establish a new reference system ( $x''$ ,  $y''$ ,  
373  $z''$ ), where the  $x''$  and  $y''$  axes are parallel to the edges at the base (the floor) of the operant  
374 box, while  $z''$  is perpendicular to the floor and passes through the central feeding needle.  
375 The position measurements returned by the head tracker can be expressed as ( $x''$ ,  $y''$ ,  $z''$ )  
376 coordinates by applying the rotation matrix  $R_{\text{box}}^{\text{cam}}$  and the translation vector  $\vec{t}_{\text{box}}^{\text{cam}}$  that map  
377 the camera reference system into this new operant box reference system.  $R_{\text{box}}^{\text{cam}}$  is in the  
378 form shown in (1), but with the angles referring to the rotation of the reference  
379 checkerboard with respect to the camera reference system;  $\vec{t}_{\text{box}}^{\text{cam}} = [\Delta t_x \quad \Delta t_y \quad \Delta t_z]$ ,  
380 where  $\Delta t$  is the distance between the origins of the two reference systems along each  
381 axis.

382 We tested the ability of the head tracker to correctly recover the Cartesian  
383 coordinates of the dots' pattern relative to the box reference system, by placing the  
384 pattern over the 5x5 grid shown in Figure 3-figure supplement 1A and collecting 30  
385 repeated, head-tracker measurements at each location. To verify the functioning of the  
386 head tracker under the same settings used for the behavioral and neurophysiological  
387 experiments (see next sections), the camera was not centered above the operant box, with  
388 the optical axis perpendicular to the floor, as previously done for the validation shown in  
389 Figure 3. Otherwise, during a neuronal recording session, the cable that connects the  
390 headstage protruding from the rat head to the preamplifier would partially occlude the  
391 camera's field of view. Hence, the need to place the camera in front of the rat, above the  
392 stimulus display, oriented with an angle of approximately  $50^\circ$  relative to the floor (see  
393 Figure 1B). This same positioning was used here to collect the set of validation  
394 measurements over the 5x5 grid. As shown in Figure 4A, the match between estimated  
395 and nominal horizontal ( $x''$ ) and vertical ( $y''$ ) coordinates of the pattern was very good  
396 (compare the red dots to the grid intersection), with a barely appreciable dispersion of the  
397 30 measurements around each nominal position value. This resulted in a good overall  
398 precision (x: RMSE =  $0.034 \pm 0.004$ ; y:  $0.158 \pm 0.010$ ) and accuracy (x: RMSE =  $0.403 \pm$   
399  $0.050$ ; y:  $1.36 \pm 0.093$ ) and of the  $x''$  and  $y''$  measurements.

400





401

402

403

404

405

406

407

408

409

410

411 **Figure 4.** Validation of the head-tracker in the reference system of the operant box. **(A)**

412 Left: validation measurements (red dots) obtained by placing the dots' pattern on a grid

413 of 5 x 5 ground-truth positions (grid's intersections; same as in Figure 3A), over the floor

414 of the operant box. Each measurement is relative to a *box reference system*, with the *x*

415 and *y* axes parallel to the floor of the operant box, and the *z* axis passing through the

416 central response port (see Figure 1A). Converting the measurements from the camera to

417 the box reference system was possible by acquiring an image of a reference

418 checkerboard, as described in the main text and shown in Figure 4-figure supplement

419 1A. The inset allows appreciating the spread of the 30 measurements taken at one of

420 the tested positions. Right: mean precision (top) and accuracy (bottom) of the head-

421 tracker measurements over the 25 tested positions (RMSE computed as in Figure 3A).

422 **(B)** Left: validation measurements (red dots) obtained by setting the yaw, pitch and roll

423 angles of the dots' pattern to a combination of 3 x 7 x 5 ground-truth values (grid's

424 intersections), using the custom assembly shown in Figure 4-figure supplement 1B.

425 Each measurement is relative to a *pose zero reference system*, obtained by acquiring an

426 image of the dots' pattern with all the angles on custom assembly set to zero. Note that,

427 given their high precision, the spread of the 30 repeated measurements taken at each

428 ground-truth value is barely appreciable. Right: mean precision (left) and accuracy (right)

429 of the head-tracker measurements over the set of tested angle combinations (RMSE

430 computed as in Figure 3A).

431

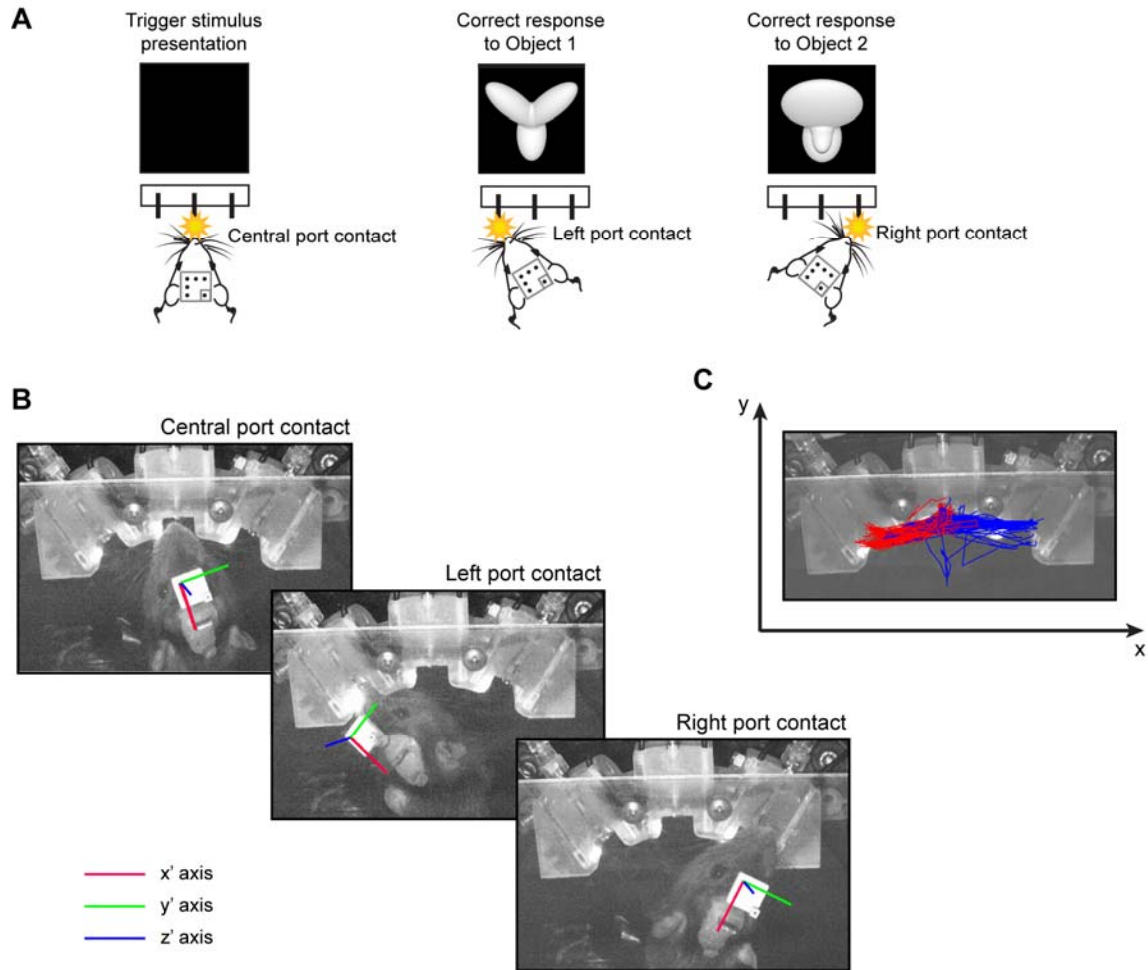
432       The Euler angles defining the pose of the head in the 3D space could also be  
433 measured relative to the operant box reference system ( $x''$ ,  $y''$ ,  $z''$ ). However, this would  
434 yield an estimate of the rotation of the dots' pattern, rather than of the rat head, in the  
435 physical environment. In fact, no matter how carefully the support holding the pattern is  
436 implanted at the time of the surgery, it is unlikely for the pattern to be perfectly aligned to  
437 the anatomical axes of the head, once put in place. In general, the  $x'$  and  $y'$  axes of the  
438 pattern reference system (see Figure 2B) will be slightly tilted, with respect to the head's  
439 anteroposterior and interaural axes. Therefore, it is more convenient to acquire an image  
440 of the pattern when the rat is placed in the stereotax, with its head parallel to the floor of  
441 the operant box, and then use such a *pose zero* of the pattern as a reference for the  
442 angular measurements returned by the head tracker. This can be achieved by defining a  
443 pose zero reference system ( $x'^0$ ,  $y'^0$ ,  $z'^0$ ) and a rotation matrix  $R_{\text{pose } 0}^{\text{cam}}$  mapping the camera  
444 reference system into this new coordinate system [the matrix is in the form shown in (1),  
445 but with the angles referring to the rotation of the pose zero of the pattern with respect to  
446 the camera reference system].

447       To test the ability of the head tracker to correctly recover the Euler angles of the  
448 dots' pattern relative to the pose zero reference system, we mounted the pattern over a  
449 custom assembly made of two 3D printed goniometers, each allowing rotations over a  
450 span of  $70^\circ$  (from  $-35^\circ$  to  $+35^\circ$ ), and a rotary stage, enabling  $360^\circ$  rotations (Figure 4-  
451 figure supplement 1B). This allowed setting the pose of the pattern over a grid of known  
452  $3 \times 7 \times 5$  combinations of yaw, pitch and roll angles. As illustrated in Figure 4B (left), we  
453 found a good match between the nominal and estimated pattern rotations (note that not all  
454  $3 \times 7 \times 5$  angle combinations were actually tested, since the dots on the pattern were not  
455 detectable at some extreme rotations). When mediated across all tested angle  
456 combinations, the resulting precision and accuracy (Figure 4B, right) were very similar to  
457 the nominal ones shown in Figure 3C (precision: roll  $0.068^\circ \pm 0.003^\circ$ , pitch  $0.076^\circ \pm$   
458  $0.004^\circ$ , yaw  $0.0409^\circ \pm 0.001^\circ$ ; accuracy: roll  $0.746^\circ \pm 0.036$ , pitch  $0.598^\circ \pm 0.048^\circ$ , yaw  
459  $0.929^\circ \pm 0.052^\circ$ ).

460

461 **Head movements of a rat performing a two-alternative forced choice task in the**  
462 **visual and auditory modalities**

463 To illustrate the application of our head-tracker *in-vivo*, we implanted a rat with an  
464 electrode array targeting hippocampus (see Materials and Methods and next section for  
465 details). The implant also included a base with a magnet that allowed attaching the dots'  
466 pattern during the behavioral/recording sessions (see Figure 2A-B). As previously  
467 explained, the animal had to interact with an array of three response ports, each equipped  
468 with a feeding needle and a proximity sensor (Figure 1). Licking the central port  
469 triggered the presentation of either a visual object (displayed on the screen placed in front  
470 of the rat) or a sound (delivered through the speakers located on the side of the monitor).  
471 Two different visual objects could be presented to the animal [same as in (Zoccolan et al.,  
472 2009; Alemi-Neissi et al., 2013a)] – Object 1 required the rat to approach and lick the left  
473 response port in order to obtain liquid reward, while Object 2 required him to lick the  
474 right response port (Figure 5A). The same applied to the two sounds, one associated to  
475 the left and the other to right response port.



476

477 **Figure 5.** Head tracking of a rat engaged in a perceptual discrimination task. (A)  
478 Illustration of the visual discrimination task. The rat learned to lick the central response  
479 port to trigger the presentation of either Object 1 or Object 2 on the stimulus display  
480 placed in front of the operant box (see Figure 1A). Presentation of Object 1 required the  
481 rat to approach and lick the response port on the left in order to correctly report its  
482 identity, while presentation of Object 2 required the animal to lick the response port on  
483 the right. (B) Example snapshots captured and processed by the head tracker at three  
484 representative times – i.e., when the rat licked the central, the left and the right response  
485 ports. The colored lines are the  $x'$  (red),  $y'$  (green) and  $z'$  (blue) axes of the reference  
486 system centered on the dots' pattern (see Figure 2C), as inferred in real time by the head-  
487 tracker (see also supplementary Figure 5-video supplement 1). (C) The trajectories  
488 followed by the nose of the rat in consecutive trials during the execution of the task are  
489 superimposed to a snapshot of the block holding the response ports, as imaged by the  
490 head tracker. The red and blue traces refer to trials in which the animal chose,

491 respectively, the left and right response port. The trajectories are plotted in the Cartesian  
492 plane corresponding to the floor of the operant box, where the  $x$  and  $y$  axes (black  
493 arrows) are, respectively, parallel and orthogonal to the stimulus display (see also Figure  
494 1A).

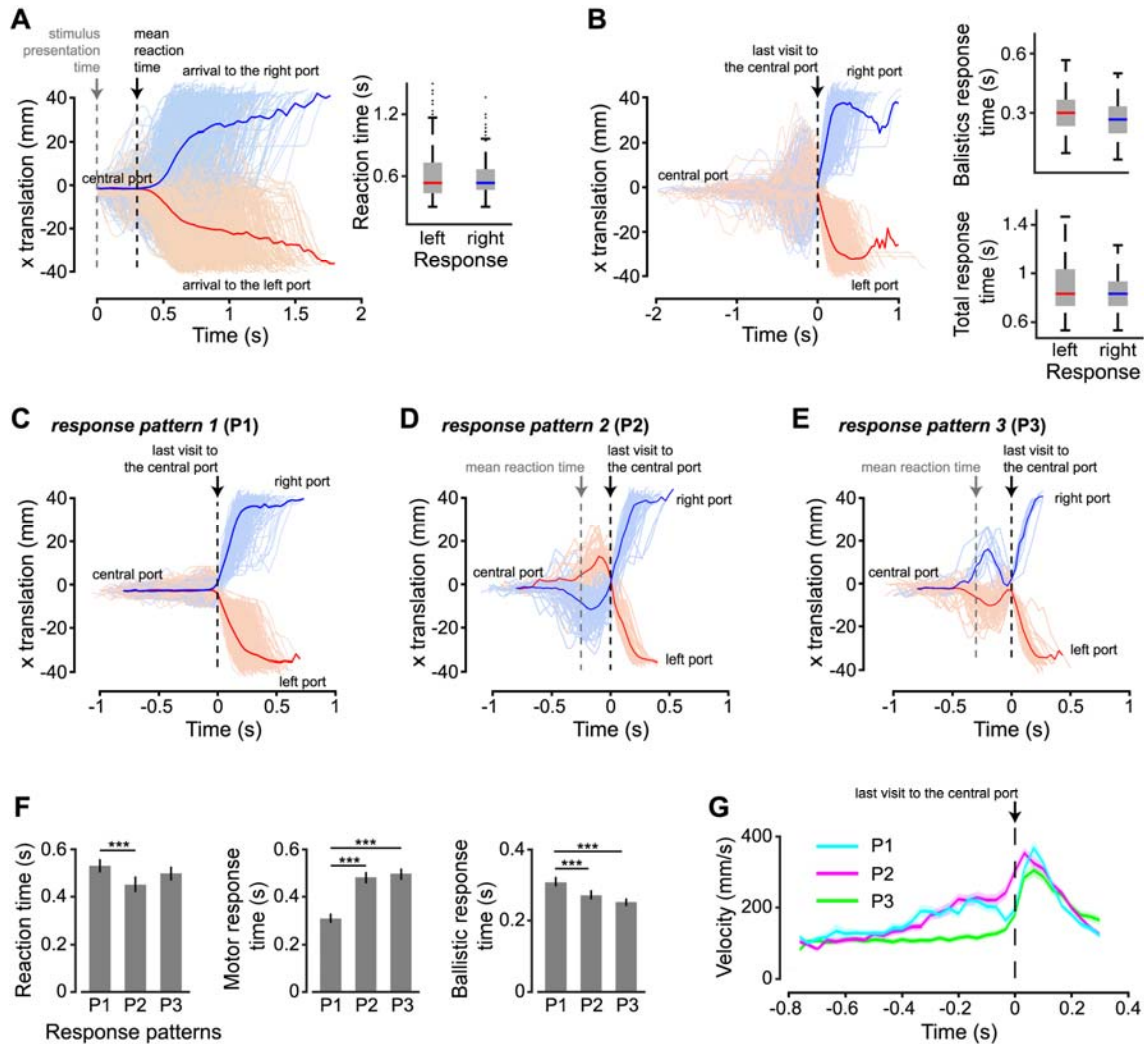
495

496 Figure 5B shows example images captured and processed by the head-tracker in  
497 three representative epochs during the execution of the task (the colored lines are the  $x'$ ,  
498  $y'$  and  $z'$  axes that define the pose of the pattern, as inferred in real time by the head-  
499 tracker; see supplementary Figure 5-video supplement 1). By tracking the position of the  
500 pattern in the 2D plane of the floor of the operant box, it was possible to plot the  
501 trajectory of the rat's nose in each individual trial of the behavioral test (Figure 5C, red  
502 vs. blue lines, referring to trials in which the animal chose, respectively, the left and right  
503 response port). The position of the nose was monitored, because it better reflects the  
504 interaction of the animal with the response sensors, as compared to the position of the  
505 pattern (the latter can be converted into the nose position by carefully measuring the  
506 distance between the origin of the pattern and the tip of the nose at the time of the  
507 surgery). With such a precise knowledge of the position and pose of the nose/head in  
508 space and time, we could address questions concerning the perceptual, motor and  
509 cognitive processes deployed by the rat during the visual and auditory discriminations,  
510 well beyond the knowledge that can be gained by merely monitoring the response times  
511 collected by the proximity sensors.

512 For example, in Figure 6A (left), we have reported the  $x$  position of the rat's nose  
513 as a function of time in all the behavioral trials collected over 14 consecutive test sessions  
514 (total of 3,413 trials). The traces are aligned to the time in which the stimulus was  
515 presented (i.e., 300 ms after the animal had licked the central port). From these traces, we  
516 computed the *reaction time* (RcT) of the rat in each trial, defined as the time, relative to  
517 the stimulus onset, in which the animal left the central sensor to start a motor response,  
518 eventually bringing its nose to reach either the left (red lines) or right (blue line) response  
519 port. As it can be appreciated by looking at the spread of both sets of traces, RcT was  
520 highly variable across trials, ranging from 300 ms to 1166 ms (extreme values not  
521 considering outliers), with a median around 530 ms (Figure 6A, right). By contrast, a

522 much smaller variability was observed for the *ballistic response time* (BRsT), i.e., the  
523 time taken by the rat to reach the left or right response port, relative to the last time he  
524 visited (i.e., passed by) the central port (Figure 6B, left), with BRsT ranging between 66  
525 ms and 566 ms (Figure 6B, right: upper box plot; median ~300 ms). This suggests that  
526 much of the variability in the *total response time* (ToRsT; i.e., the time taken to reach the  
527 left or right response port, relative to the stimulus onset; Figure 6B, right: bottom box  
528 plot) has to be attributed to the perceptual/decisional process required for the correct  
529 identification of the stimulus. However, a close inspection of the response trajectories of  
530 Figure 6B (aligned to the time in which the ballistic motor response was initiated) shows  
531 that the rat, after leaving the central port in response to the stimulus, often did not point  
532 directly to the port that he would eventually choose (referred to as the *selected response*  
533 *port* in what follows). In many trials, the final ballistic response was preceded by earlier  
534 movements of the head, either towards the selected response port or the opposite one  
535 (referred to as the *opposite response port* in what follows).

536



537

538 **Figure 6.** Statistical characterization of the head's displacements performed by the rat  
 539 during the perceptual discrimination task. **(A)** Left: position of the rat's nose as a function  
 540 of time, along the x axis of the Cartesian plane corresponding to the floor of the operant  
 541 box (i.e., x axis in Figure 5C). The traces, recorded in 3,413 trials over the course of 14  
 542 consecutive sessions, are aligned to the time in which the stimulus was presented (gray  
 543 dashed line). The black dashed line indicates the mean reaction time (as defined in the  
 544 main text). The red and blue colors indicate trials in which the rat chose, respectively,  
 545 the left and right response port. Right: box plot showing the distributions of reaction  
 546 times for the two classes of left and right responses. **(B)** Left: same trajectories as in **A**,  
 547 but aligned to the last time the rat visited the central port (dashed line). Right: box plots  
 548 showing the distributions of ballistic response times (top) and total response times  
 549 (bottom) for the two classes of left and right responses (see main text for definitions). **(C)**

550 Subset of the response patterns (referred to as *P1*) shown in **A** and **B**, in which the rat,  
551 after leaving the central port, made a direct ballistic movement to the selected response  
552 port (traces' alignment as in **B**). (**D**) Subset of the response patterns (referred to as *P2*)  
553 shown in **A** and **B**, in which the rat, before making a ballistic movement towards the  
554 selected response port, made an initial movement towards the opposite port (traces'  
555 alignment as in **B**). (**E**) Subset of the response patterns (referred to as *P3*) shown in **A**  
556 and **B**, in which the rat made an initial movement towards the selected response port,  
557 then moved back to the central port, and finally made a ballistic moment to reach the  
558 selected port (traces' alignment as in **B**). (**F**) Comparisons among the mean reaction  
559 times (left), among the mean motor response times (middle; see main text for a  
560 definition) and among the mean ballistic response times (right) that were measured for  
561 the three types of trajectories (i.e., *P1*, *P2* and *P3*) shown, respectively, in **C**, **D** and **E**  
562 (\*\*\*)  $p < 0.001$ ; two-tailed, unpaired t-test). Error bars are SE of the mean. (**G**) Velocity of  
563 the rat's nose as a function of time for the three types of trajectories *P1*, *P2* and *P3*.

564

565 To better investigate rat response patterns, we separated the recorded trajectories  
566 into three response patterns. In the first response pattern (*P1*), the rat, after leaving the  
567 central port, made a direct ballistic movement to the selected response port (either left or  
568 right; Figure 6C). In the second response pattern (*P2*), the rat made an initial movement  
569 towards the opposite response port, before correcting himself and making a ballistic  
570 movement towards the selected response port (Figure 6D). In the third response pattern  
571 (*P3*), the rat made an initial movement towards the selected response port, but then  
572 moved back to the central port, before approaching again, with a ballistic moment, the  
573 selected port (Figure 6E). Interestingly, RcT was significantly smaller in *P2* than in *P1* ( $p$   
574  $< 0.001$ ; two-tailed, unpaired t-test; Figure 6F, leftmost bar plot), suggesting that the  
575 trials in which the animal reversed his initial decision were those in which he made an  
576 “impulsive” choice that he eventually corrected. As expected, the *motor response time*  
577 (MoRsT; i.e., the time taken by the animal to reach the selected response port, after  
578 leaving, for the first time, the central port) was substantially lower in *P1*, as compared to  
579 *P2* and *P3*, given the indirect trajectories that the latter trial types implied ( $p < 0.001$ ;  
580 two-tailed, unpaired t-test; Figure 6F, central bar plot). By contrast, BaRsT was slightly,  
581 but significantly higher in *P1* than in *P2* and *P3*, indicating that ballistic movements were

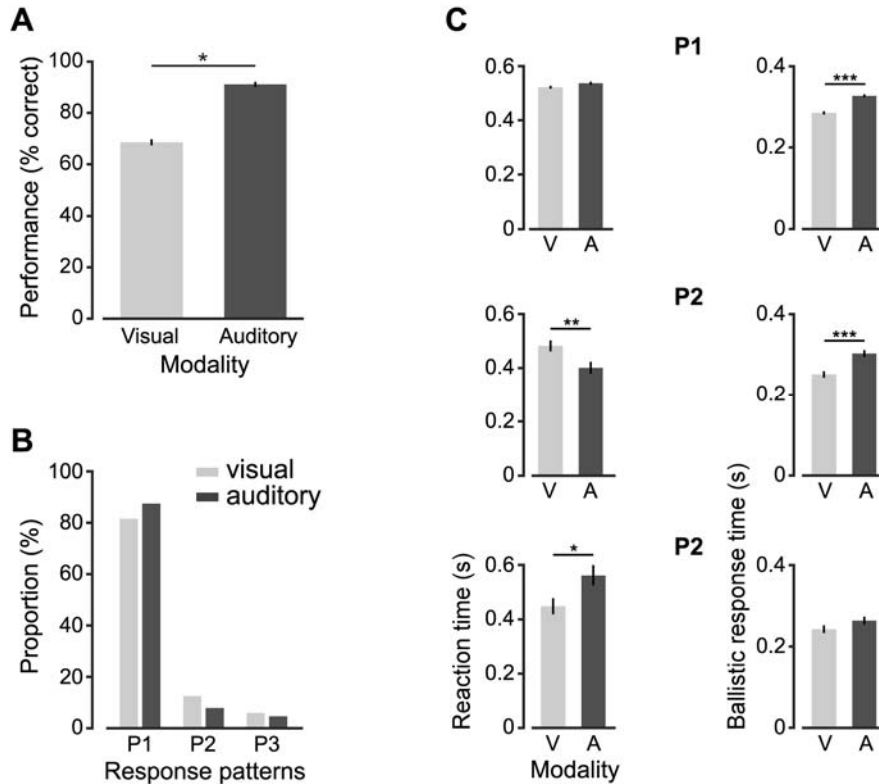


582 faster, when they followed a previously aborted choice ( $p < 0.001$ ; two-tailed, unpaired t-  
583 test; Figure 6F, rightmost bar plot).

584 To better understand this phenomenon, we plotted the average velocity of the rat's  
585 nose as a function of time for the three response patterns (Figure 6G; curves are aligned  
586 to the onset of the ballistic motor response; dashed line). As expected, in P1 (green  
587 curve), the velocity was close to zero till the time in which the ballistic response was  
588 initiated. By contrast, in P2 (magenta curve), the velocity was already high at the onset of  
589 the ballistic movement, because the animal's nose passed by the central sensor when  
590 sweeping from the opposite response port to the selected one (see Figure 6D). This  
591 momentum of the rat's head was thus at the origin of the faster ballistic responses in P2,  
592 as compared to P1. In the case of P3 (cyan curve), there was no appreciable difference of  
593 velocity with P1 at the onset of the ballistic movements. This is expected, given that the  
594 rat moved twice in the direction of the selected port and, in-between the two actions, he  
595 stopped at the central sensor (see Figure 6E). However, following the onset of the  
596 ballistic movement, the rat reached a larger peak velocity in P3 than in P1, which  
597 explains the shorter time needed to complete the ballistic response in the former trials'  
598 type. This may possibly indicate a larger confidence of the rat in his final choice,  
599 following an earlier, identical response choice that was later aborted.

600 We also used the head tracker to inquire whether the rat deployed different  
601 response/motor patterns depending on the sensory modality of the stimuli he had to  
602 discriminate. Rat performance was higher in the sound discrimination task, as compared  
603 to the visual object discrimination task ( $p < 0.001$ ; two-tailed, unpaired t-test; Figure 7A).  
604 Consistently with this observation, the fraction of trials in which the animal aborted an  
605 initial perceptual choice (i.e., response patterns P2 and P3), as opposed to make a direct  
606 response to the selected port (i.e., response pattern P1), was significantly larger in visual  
607 than in auditory trials ( $p < 0.01$ ,  $\chi^2$  test for homogeneity; Figure 7B). This means that the  
608 animal was less certain about his initial decision in the visual trials, displaying a tendency  
609 to correct more often such decision, as compared to the auditory trials. Interestingly, the  
610 lower perceptual discriminability of the visual stimuli did not translate into a general  
611 tendency of reaction times to be longer in visual than auditory trials. When the animal  
612 aimed directly to the selected port (P1; the vast majority of trials, as it can be appreciated

613 in Figure 7B), no significant difference in RcT was observed between visual and auditory  
614 trials ( $p > 0.05$ ; two-tailed, unpaired t-test; Figure 7C, top-left bar plot). By contrast, in  
615 trials in which the rat corrected his initial decision (P2), RcT was significantly longer in  
616 visual than auditory trials ( $p < 0.01$ ; two-tailed, unpaired t-test; Figure 7C, middle-left bar  
617 plot), but the opposite trend was found in trials in which the animal swung back and forth  
618 to the selected response port (P3;  $p < 0.05$ ; two-tailed, unpaired t-test; Figure 7C, bottom-  
619 left bar plot). We found instead a general tendency of the rat to make faster ballistic  
620 responses in visual than auditory trials, with this trend being significant in P1 and P2 ( $p <$   
621  $0.001$ ; two-tailed, unpaired t-test; Figure 7D, top-right and middle-right bar plots). It is  
622 hard to interpret these findings, which could indicate a larger impulsivity of the animal in  
623 the visual discrimination, but, possibly, also a larger confidence in his decision.  
624 Addressing more in depth this issue is obviously beyond the scope of our study, since it  
625 would require measuring the relevant metrics (e.g., RcT and BaRsT) over a cohort of  
626 animals, while our goal here was simply to provide an in-vivo demonstration of the  
627 working principle of our head-tracker and suggest possible ways of using it to investigate  
628 the perceptual, decision and motor processes involved in a perceptual discrimination task.  
629



630

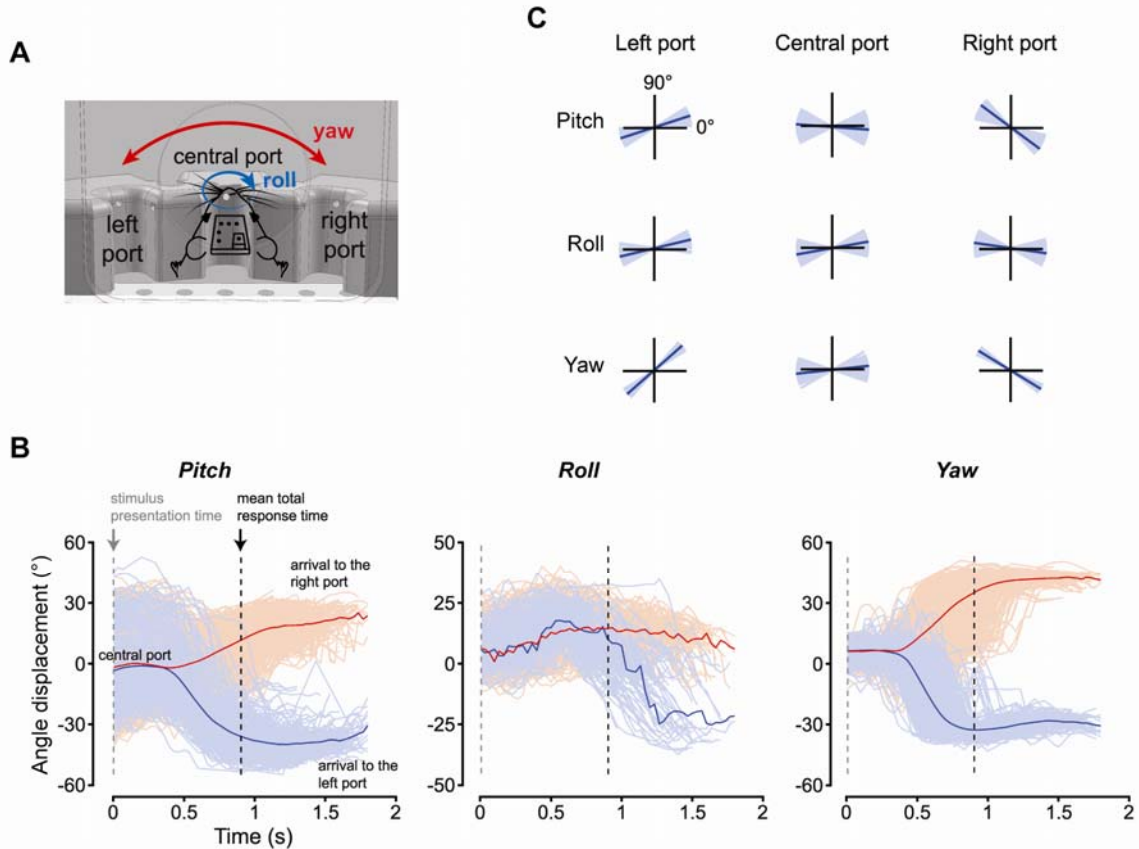
631 **Figure 7.** Statistical comparison of the head's displacements performed by the rat in  
 632 visual and auditory trials. **(A)** Comparison between the performances attained by the rat  
 633 in the visual (light gray) and auditory (dark gray) discrimination tasks (\*\*\*)  $p < 0.001$ ; two-  
 634 tailed, unpaired t-test). **(B)** Proportion of response patterns of type P1, P2 and P3 (see  
 635 Figure 6C-E) observed across the visual (light gray) and auditory (dark gray) trials. The  
 636 two distributions were significantly different ( $p < 0.01$ ,  $\chi^2$  test). **(C)** Comparisons between  
 637 the mean reaction times (left) and between the mean ballistic response times (right) that  
 638 were measured in visual (light gray) and auditory (dark gray) trials for each type of  
 639 trajectories (i.e., P1, P2 and P3). \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; two-tailed,  
 640 unpaired t-test.

641

642 As a further example of the kind of behavioral information that can be extracted  
 643 using the head-tracker, we analyzed the pose of the rat's head during the execution of the  
 644 task (Figure 8). As shown in Figure 8B, at the time the rat triggered the stimulus  
 645 presentation (time 0), his head was, on average across trials (thick curves), parallel to the  
 646 floor of the operant box (i.e., with pitch  $0^\circ$  and roll close to  $0^\circ$ ) and facing frontally the

647 stimulus display (i.e., with yaw close to 0°). However, at the level of single of trials (thin  
648 curves), the pose of the head was quite variable, with approximately a  $\pm 30^\circ$  excursion in  
649 the pitch, and a  $\pm 15^\circ/20^\circ$  excursion in the roll and yaw. This can be also appreciated by  
650 looking at the polar plots of Figure 8C (central column), which report the average pitch,  
651 roll and yaw angles that were measured over the 10 frames (~300 ms) following the  
652 activation of each response port (thin lines: individual trials; thick lines: trials' average).  
653 Being aware of such variability is especially important in behavioral and  
654 neurophysiological studies of rodent visual perception (Zoccolan, 2015). In fact, a  
655 variable pose of the head at the time of stimulus presentation implies that the animal  
656 viewed the stimuli under quite different angles across repeated behavioral trials. This, in  
657 turns, means that the rat had to deal with a level of variation in the appearance of the  
658 stimuli on his retina that was larger than that imposed, by the design, by the  
659 experimenter. In behavioral experiments where the invariance of rat visual perception is  
660 under investigation, this is not an issue, because, as observed in (Alemi-Neissi et al.,  
661 2013), it can lead at most to an underestimation (not to an overestimation) of rat invariant  
662 shape-processing abilities. However, in studies were a tight control over the retinal image  
663 of the visual stimuli is required, the trial-by-trial variability reported in Figure 8B-C  
664 indicates that the use of a head-tracker is necessary to measure, and possibly compensate,  
665 the change of viewing angle occurring across repeated stimulus presentations. This  
666 applies, for instance, to neurophysiological studies of visual representations in  
667 unrestrained (i.e., not head-fixed) rodents, especially when localized stimuli (e.g., visual  
668 objects) are used to probe low- and middle-level visual areas. For instance, head-tracking,  
669 ideally also paired with eye-tracking, would be necessary to investigate putative ventral  
670 stream areas in unrestrained rats, as recently done in anesthetized (Tafazoli et al., 2017;  
671 Matteucci et al., 2019) or awake, but head-fixed animals (Vinken et al., 2014, 2016,  
672 2017; Kaliukhovich and Op de Beeck, 2018). By contrast, the issue of pose variability  
673 afflicts less neurophysiological studies targeting higher-order association or decision  
674 areas, especially when using large (ideally full-field) periodic visual patterns (e.g.,  
675 gratings) (Nikbakht et al., 2018).

676



677

678 **Figure 8.** Statistical characterization of the head's rotations performed by the rat during  
679 the perceptual discrimination task. **(A)** Illustration of the yaw and roll rotations that the  
680 rat's head can perform, relative to its reference pose (see main text). **(B)** Time course of  
681 the pitch (left), roll (middle) and yaw (right) angles during repeated trials of the  
682 perceptual discrimination task. The red and blue colors indicate trials in which the rat  
683 chose, respectively, the left and right response port. Traces are aligned to the time in  
684 which the stimulus was presented (dashed gray line). The mean total response time (i.e.,  
685 the mean time of arrival to the selected response port) is shown by the dashed black  
686 line. **(C)** The pitch (top), roll (middle) and yaw (bottom) angles measured by the head  
687 tracker in individual trials (thin blue lines) and, on average, across all trials (thick black  
688 lines), at the times the rat reached the left, central and right response ports.

689

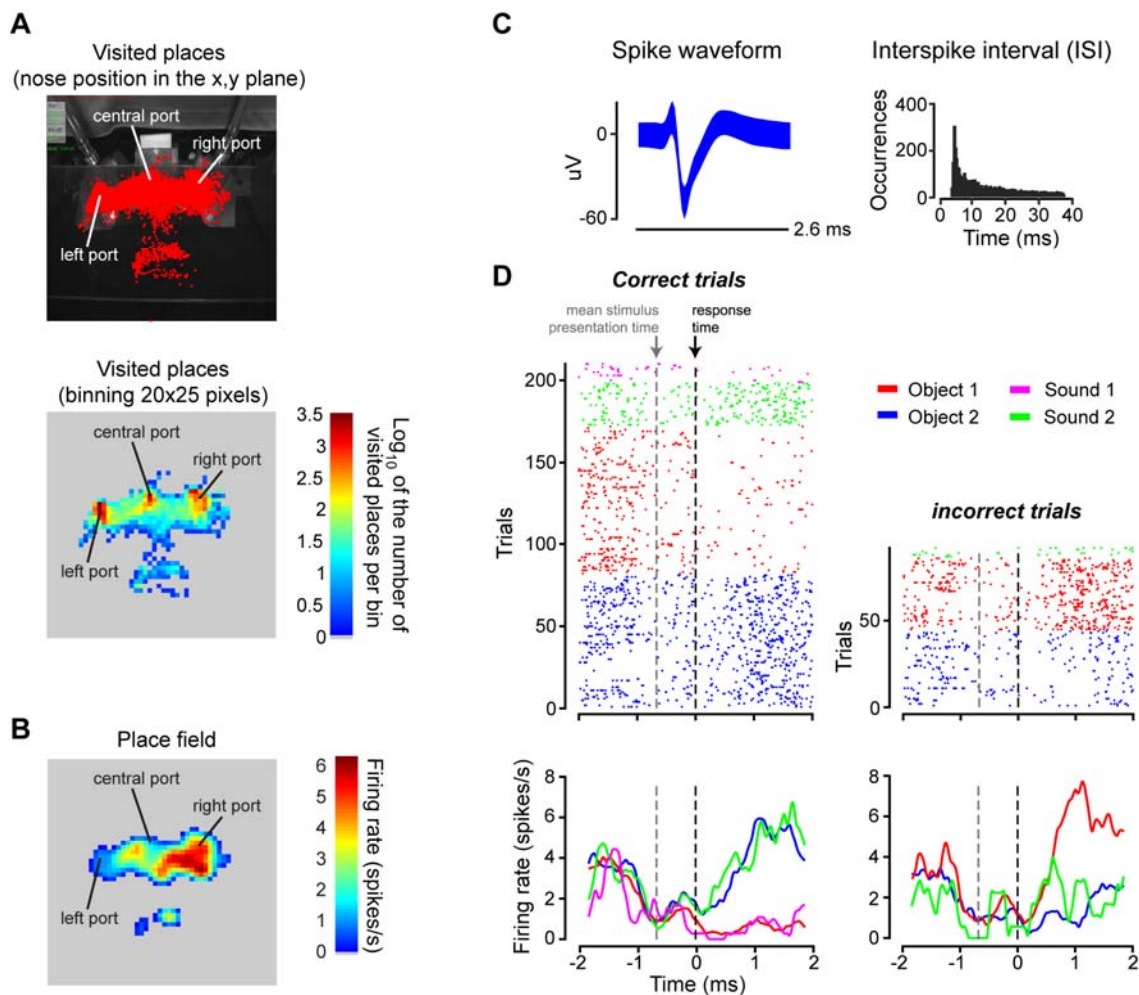
690 Monitoring the pose of the head during the behavioral trials also revealed that the  
691 rat approached the lateral responses ports with his head at port-specific angles (compare  
692 red vs. blue curves in Figure 8B, and left vs. right columns in Figure 8C). Unsurprisingly,

693 the yaw angle was the opposite for left (about 40°) and right (about -35°) responses, since  
694 the animal had to rotate his head towards opposite direction to reach the lateral response  
695 ports (see Figure 8A, red arrows). It was less obvious to observe opposite rotations also  
696 about the roll and pitch axes, which indicate that the rat bent his head in port-specific  
697 ways to reach each response port and lick from the corresponding feeding needle. Again,  
698 this information is relevant for behavioral studies of rodent visual perception, where the  
699 stimulus is often left on the screen after the animal makes a perceptual choice and during  
700 the time he retrieves the liquid reward (Zoccolan et al., 2009; Alemi-Neissi et al., 2013;  
701 Rosselli et al., 2015; Djurdjevic et al., 2018). This implies that the animal experiences  
702 each stimulus from a port-specific (and therefore stimulus-specific) viewing angle for a  
703 few seconds after the choice. As pointed out in (Djurdjevic et al., 2018), this can explain  
704 why rats learn to select specific stimulus features to process a given visual object.

#### 705 **Simultaneous head tracking and neuronal recordings during a two-alternative** 706 **forced choice discrimination task**

707 To illustrate how our head tracker can be combined with the recording of neuronal  
708 signals, we monitored the head movements of the rat during the execution of the  
709 visual/auditory discrimination task, while recording the activity of hippocampal neurons  
710 in CA1. Given that, in rodents, hippocampal neurons often code the position of the  
711 animal in the environment (Moser et al., 2008, 2015), we first built a map of the places  
712 visited by the rat while performing the task (Figure 9A, top). It should be noticed that,  
713 differently from typical hippocampal studies, where the rodent is allowed to freely move  
714 inside an arena, the body of the rat in our experiment was at rest, while his head, after  
715 entering through the viewing hole, made small sweeps among the three response ports.  
716 Therefore, the map of visited positions shown in Figure 9A refers to the positions of the  
717 rat's nose, rather than to his body. Thus, by binning the area around the response ports,  
718 we obtained a density map of visited locations by the nose (Figure 9A, bottom). Not  
719 surprisingly, this map revealed that the rat spent most of the time with his nose very close  
720 to one of the response ports (red spots). Figure 9B shows instead the average firing rate  
721 of an example, well-isolated hippocampal neuron (whose waveforms and inter-spike  
722 interval distribution are shown in Figure 9C) as a function of the position of the rat's nose

723 (only locations that were visited more than 10 times were considered). The resulting  
724 spatial map displayed a marked tendency of the neuron to preferentially fire when the rat  
725 approached the right response port. In other words, this neuron showed the sharp spatial  
726 tuning that is typical of hippocampal place cells, but over a much smaller spatial extent  
727 (i.e., over the span of his head movements around the response ports) than typically  
728 observed in freely moving rodents.  
729



730

731 **Figure 9.** Place field of a rat hippocampal neuron during the execution of the perceptual  
732 discrimination task. **(A)** Top: map of the places visited by the nose of the rat (red dots)  
733 around the area of the response ports, while the animal was performing the  
734 discrimination task. Bottom: density map of visited locations around the response ports.  
735 The map was obtained by: 1) dividing the image plane in spatial bins of 20x25 pixels; 2)  
736 counting how many times the rat visited each bin; and 3) taking the logarithm of the

737 resulting number of visits per bin (so as to allow a better visualization of the density  
738 map). **(B)** Place field of a well-isolated hippocampal neuron (as shown in **C**). The  
739 average firing rate of the neuron (i.e., the number of spikes fired in a 33 ms time bin,  
740 corresponding to the duration of a frame captured by the head tracker) was computed  
741 across all the visits the rat's nose made to any given spatial bin around the response  
742 ports (same binning as in **B**). Only bins with at least 10 visits were considered, and the  
743 raw place field was smoothed with a Gaussian kernel with  $\sigma = 1.2$  bins. **(C)**  
744 Superimposed waveforms (left) and inter-spike interval distribution (right) of the recorded  
745 hippocampal neuron. **(D)** Time course of the activity of the neuron, during the execution  
746 of the task. The dots in the raster plots (top) indicate the times at which the neuron fired  
747 an action potential (or spike). The peri-response time histograms (PRTs) shown in the  
748 bottom were obtained from the raster plots by computing the average number of spikes  
749 fired by the neuron across repeated trials in consecutive time bins of 0.48 ms. The color  
750 indicates the specific stimulus condition that was presented in a given trial (see caption  
751 in the figure). The panels on the left refer to trials in which the rat's choice was correct,  
752 while the panels on the right refer to trials in which his response was incorrect. Both the  
753 raster plots and the PRTs are aligned to the time of arrival to the response port (time 0;  
754 black dashed line). The gray dashed line shows the mean stimulus presentation time  
755 (gray dashed line), relative to the time of the response.

756

757 To verify that the neuron mainly coded positional information, we plotted its  
758 activity as a function of time, across the various epochs of the discrimination task, as well  
759 as a function of the sensory stimuli the rat had to discriminate and of the correctness of  
760 his choices. This is illustrated by the raster plots of Figure 9D (top), where each dot  
761 shows the time at which the neuron fired an action potential, before and after the animal  
762 made a perceptual choice (i.e., before and after he licked one of the lateral response  
763 ports). Individual rows show the firing patterns in repeated trials during the task, with the  
764 color coding the identity of the stimulus (see caption). In addition, the trials are grouped  
765 according to whether the rat's response was correct (left) or not (right). Each raster plot  
766 was then used to obtain the peri-response time histograms (PRTs) shown in the bottom  
767 of Figure 9D, where the average firing rate of the neuron across trials of the same kind is  
768 reported as a function of time. As expected for a place cell, and consistently with the  
769 spatial tuning shown in Figure 9B, the neuron started to respond vigorously after the rat



770 licked the right response port and did so regardless of whether the stimulus was auditory  
771 or visual (see, respectively, the green and blue dots/lines in Figure 9D). In addition, the  
772 neuron also fired when the rat licked the right port in response to the visual stimulus  
773 (Object 1) that required a response to be delivered on the left port, i.e., on trials in which  
774 his choice was incorrect (red dots/line).

775 Obviously, our analysis of this example neuronal response pattern is far from  
776 being systematic. It is simply meant to provide an illustration of how our light-weight,  
777 portable head-tracking system can be applied to study the motor/behavioral patterns of  
778 small rodents (and their neuronal underpinnings) in experimental contexts where the  
779 animals do not navigate large environments but are confined to a restricted (often small)  
780 span of spatial locations.

781

## 782 **Materials and methods**

### 783 **Experimental rig for behavioral tests and in-vivo electrophysiology**

784 The rig used to administer to the rat the two-alternative forced-choice (2AFC)  
785 discrimination task (Zoccolan, 2015; Zoccolan and Di Filippo, 2018) has been already  
786 described in the Results. Briefly (with reference to Figure 1 A and B), a custom-made  
787 operant box was designed with the CAD software SolidWorks (Dassault Systèmes) and  
788 then built using black and transparent Plexiglas. The box was equipped with a 42-inch  
789 LCD monitor (Sharp, PN-E421) for presentation of visual stimuli and an array of three  
790 stainless steel feeding needles (Cadence Science), 10 mm apart from each other  
791 connected to three proximity sensors. The left and right feeding needles were connected  
792 to two computer-controlled syringe pumps (New Era Pump Systems NE-500), for  
793 automatic pear juice delivery. Each feeding needle was flanked by a light-emitting diode  
794 (LED) on one side and a photodiode on the other side, so that, when the rat licked the  
795 needle, he broke the light beam extending from the LED to the photodiode and his  
796 responses was recorded. The front wall of the operant box had a rectangular aperture,  
797 which was 4-cm wide and extended vertically for the whole height of the wall, so as to  
798 allow room for the cables connecting the implanted electrode array to the preamplifiers of  
799 the acquisition system. The rat learned to insert his head through this aperture, so as to

800 reach the feeding needles and face the stimulus display, which was located 30 cm apart  
801 from his nose. Two speakers positioned at the sides of the monitor were used for playing  
802 the sound stimuli. Stimulus presentation, input and output devices, as well as all task-  
803 relevant parameters and behavioral data acquisition were controlled with the freeware,  
804 open-source software MWorks (<https://mworks.github.io/>) running on a Mac Mini  
805 (Apple; solid cyan and black arrows in Figure 1A).

806 During the neurophysiological experiments, stimulus presentation and collection  
807 of behavioral responses were synchronized with the amplification and acquisition of  
808 extracellular neuronal signals from hippocampus (red arrow in Figure 1A), performed  
809 using a system three workstation (Tucker-Davis Technologies – TDT), with a sampling  
810 rate of 25 kHz, running on a Dell PC. Specifically, MWorks sent a code with the identity  
811 and time of the stimulus presented in every trial to the TDT system via a UDP connection  
812 (dotted black arrow). The TDT system also controlled the acquisition of the frames by the  
813 camera of the head tracker, by generating a square wave that triggered the camera image  
814 uptake (dashed black arrow). The camera, in turn, generated a unique identification code  
815 for every acquired image. This code was saved by the PC running the head-tracking  
816 software (along with the Cartesian coordinates and pose of the head) and also fed back to  
817 the TDT (dashed green arrow), which saved it in a data file along with the  
818 neurophysiological and behavioral recordings. The square wave had a fixed period and  
819 duty cycle, but it was adjustable by the user from the TDT graphical interface. In our  
820 experiments, the period was 23 ms (about 50 Hz) and the duty cycle was around 33%.

### 821 **Perceptual discrimination tasks, surgery and neuronal recordings**

822 One adult male Long Evans rat (Charles rivers Laboratories) was used for the validation  
823 of the head tracking system. The animal was housed in a ventilated cabinet (temperature  
824 controlled) and maintained on a 10/14-h light/dark cycle. The rat weighed approximately  
825 250 gr at the onset of the experiment and grew to over 500 gr at the end of the study. The  
826 rat received a solution of water and pear juice (ratio 1:5) as a reward during each training  
827 session and, in addition, he had access to water ad libitum for 1 h after the training. All  
828 animal procedures were in agreement with international and institutional standards for the  
829 care and use of animals in research and were approved by the Italian Ministry of Health:

830 project N. DGSAF 22791-A, submitted on Sep. 7, 2015 and approved on Dec. 10, 2015  
831 (approval N. 1254/ 2015-PR).

832 The rat was trained in a visual and sound recognition task in the rig described in  
833 the previous section. The task required the animal to discriminate between either two  
834 visual objects or two sounds. Each trial started when the animal touched the central  
835 response port, which triggered the stimulus presentation. After the presentation of the  
836 stimulus on the monitor or the playback of the sound, the rat had to report its identity by  
837 licking either the left or the right response port. Each stimulus was associated to one  
838 specific port. Hence, only one action, either licking the left or right port, was associated  
839 to the reward in any given trial. Correct object identification was followed by the delivery  
840 of the pear juice-water solution, while incorrect response yielded a 1-3 s time out, with no  
841 reward delivery and a failure tone played along with the flickering of the monitor from  
842 black to middle grey at 15 Hz. The stimuli were presented for 1 sec or until the animal  
843 licked one of the lateral response ports, independently of the correctness of the choice.

844 The visual stimuli consisted of a pair of three-lobed objects, previously used by  
845 our group in several studies of visual perceptual discrimination in rats (Zoccolan et al.,  
846 2009; Alemi-Neissi et al., 2013a; Rosselli et al., 2015a; Djurdjevic et al., 2018).  
847 Specifically, each object was a rendering of a three-dimensional model built using the ray  
848 tracer POV-Ray (<http://www.povray.org>). The sound stimuli were two pure tones, at 600  
849 Hz and a 6000 Hz respectively (sound level: 55 dB approximately). Sounds were  
850 delivered from two speakers located symmetrically on the both sides of the front part of  
851 the operant box, so that the sound level at the position of the animal's ears was equal  
852 when the animal's nose was near the central response port at the onset of the trial.

853 Once rat reached  $\geq 70$  % correct behavioral performance, it was implanted with  
854 an electrode array for chronic recordings. To this aim, the animal was anaesthetized with  
855 isofluorane and positioned in a stereotaxic apparatus (Narishige, SR-5R). A craniotomy  
856 was made above the dorsal hippocampus and a 32-channel Zif-clip array (Tucker-Davis  
857 Technologies Inc., TDT) was lowered into the craniotomy. Five stainless steel screws  
858 were inserted into the skull (three anterior, one lateral and two posterior to the  
859 craniotomy) in order to give anchoring to the implant cementation. Around the implant,

860 we put Haemostatic gelatin sponge (Spongostan™ dental, Ethicon, Inc) saturated with  
861 sterile sodium chloride solution to protect the brain from dehydration, and then silicon  
862 (Kwik-Cast™, World Precision Inst) to seal the craniotomy and protect it from the dental  
863 cement (Secure, Sun Medical Co LTD) that was finally used to secure the whole implant  
864 to the skull. Hippocampal stereotaxic coordinates were: -3.7 mm AP, -3.5 mm ML. The  
865 final depth of the electrodes to target CA1 was around -2.2mm, and for the CA3  
866 subregion was around -3.4mm. In order to place over the head the rat the dots' pattern  
867 that was necessary for head tracking (see next sections), a complementary connector to  
868 the one mounted on the pattern was also cemented on the head, anterior to the electrode  
869 array (with a magnet on top; 210 g/cm<sup>2</sup> holding force; Figure 2A-B). The rat was given  
870 antibiotic enrofloxacin (Baytril; 5 mg/kg) and caprofen (Rimadyl; 2.5 mg/kg,  
871 subcutaneous injection) for prophylaxis against infections and postoperative analgesia for  
872 the next three days post-surgery. The animal was allowed to recover for 7 to 10 days after  
873 the surgery, during which he had access to water and food ad libitum. The behavioral and  
874 recording sessions in the operant box were resumed after this recovery period. Action  
875 potentials (spikes) in the extracellular signals acquired by the TDT system were detected  
876 and sorted for each recording site separately, using Wave Clus (Quiroga et al., 2004) in  
877 Matlab (The MathWorks). Online visual inspection of prominent theta waveforms in  
878 addition to histology confirmed the position of the electrodes.

### 879 **Head-tracking system**

880 The head tracker (Figure 1) consists of an industrial monochromatic CMOS camera  
881 (Point gray FLEA model), a far-red illuminator, a dedicated PC (Intel Xeon HP  
882 Workstation Z620 with Xeon CPU 2.5GHz, and RAM 16GB), and a three-dimensional  
883 pattern of dots, mounted over the head of the animal and imaged by the overhead camera  
884 (Figure 2A). The illuminator was made of a matrix of 4x4 LEDs, with dominant  
885 wavelength at 730 nm (OSLON SSL 150, PowerCluster LED Arrays) and a radiance  
886 angle of [-40°,40°], and was powered at 100-150 mW. It was mounted right above the  
887 stimulus display, oriented towards the operant box with an angle of approximately 50°  
888 with respect to the display (Figure 1B). The camera was set in the external trigger mode  
889 and the triggers it received from the TDT system (dashed black arrow in Figure 1A) were

890 counted, so that each frame was assigned a unique code, which was encoded in the first  
891 four pixels of each acquired image. The same codes were sent to the TDT (dashed green  
892 arrow), which stored them, so as to allow the image frames (and, therefore, the positional  
893 and pose information returned by the head tracker) to be aligned to the  
894 neurophysiological and behavioral data. In our validation analyses and in our tests with  
895 the implanted rat, the CMOS sensor integration time was set at 3 ms. This value was  
896 chosen after preliminary tests with the implanted rat performing the discrimination task,  
897 in such a way to guarantee that the images of the dots' pattern acquired during fast  
898 sweeps among the response ports were not blurred. The camera was placed above the  
899 stimulus display, oriented towards the operant box with an angle of approximately 50°  
900 with respect to the display (Figure 1B), although some of the validation measures  
901 presented in the Results (Figure 3) were obtained with the camera mounted in such a way  
902 to have its optical axis perpendicular to the floor of the measurements' area.

903 Obviously, following the generation of a trigger signal by the TDT, a propagation  
904 delay occurred before the camera started integrating for the duration of the exposure time.  
905 This delay was about 5  $\mu$ sec and has been measured as suggested by the camera  
906 constructor. That is, we configured one of the camera's GPIO pins to output a strobe  
907 pulse and we connected to an oscilloscope both the input trigger pin and the output strobe  
908 pin. The time delay from the trigger input to the complete image formation was then  
909 given by the sum of the propagation delay and exposure time, where the latter, as  
910 explained above, was fixed at 3 ms. Any other possible time delays like the sensor  
911 readout and the data transfer toward the PC, do not affect real-time acquisition, since the  
912 content of the image and its frame number is entirely established at the end of the  
913 exposure time

914 After the image has been read from the camera memory buffer, it becomes  
915 available to the head-tracking software (see next section) and, before any image  
916 processing occurs, the embedded frame number is extracted from the image and sent back  
917 to the TDT system via UDP protocol. Due to the data transfer over the USB 3.0 and  
918 UDP connections, the frame number information reaches the TDT system with an  
919 unpredictable time distance from the signal that triggered the acquisition of the frame  
920 itself. However, synchronizing the triggering pulses and the frame-numbers is

921 straightforward when considering the frame numbers and the number of trigger pulses  
922 generated. To validate our synchronization method, we conducted a double check by  
923 comparing the synch by frame-number with the synch by timestamps. The second  
924 method uses the timestamp of the frame-number received by TDT and looks for the  
925 trigger pulse that could have generated the frame in an interval of time [-0.0855 -0.0170]  
926 msec before the frame number arrived. In particular, it assigns to the current frame-  
927 number the more recent pulse that has not yet been assigned before. The two methods  
928 gave identical results over long sessions.

929 As explained in the Results, the head tracker works by imaging a 3D pattern  
930 (shown in Figure 2A), consisting of 5 coplanar black dots over a white background  
931 (approximately 3.5 mm apart from each other), plus a sixth dot located over an elevated  
932 pillar (approximately 5 mm tall). The 5 coplanar dots were arranged in a L-like shape,  
933 while the elevated dot was placed at the opposite side with respect to the corner of the L  
934 shape, so as to minimize the chance of occluding the other dots. The 3D structure holding  
935 the dots was designed with the CAD software SolidWorks (Dassault Systèmes) and then  
936 3D printed. The dots were printed on a standard white paper sheet, which was then cut to  
937 the proper size and glued over the 3D printed structure. This structure also included an  
938 arm, with a magnet at the bottom, which allowed the pattern to be mounted over an  
939 apposite support (holding a complementary magnet) that was surgically implanted on the  
940 rat's head (Figure 2A-B).

941 Since the head tracking algorithm (see next sections) requires a precise  
942 knowledge of the spatial arrangement of the dots relative to each other, the distance  
943 between each pair of dots, as well as the height of the pillar, were carefully measured  
944 using a caliper and by acquiring high-magnification images of the pattern along with a  
945 precision ruler placed nearby (SE® resolution 0.5 mm). This information is inserted in an  
946 apposite configuration file and is part of the calibration data that are required to operate  
947 the head tracker. The other calibration information is the internal camera  
948 parameters  $K$ , which have to be pre-computed by the camera calibration procedure (see  
949 apposite section below). Once this information is known, it can be stored and then  
950 loaded from the configuration file at the beginning of each behavioral/recording session,

951 provided that the position of the camera in the rig is not changed and the same dots'  
952 pattern is always used.

### 953 **Overview of the head-tracking algorithm**

954 Our head-tracking algorithm consists of three different functional parts: 1) the *Point*  
955 *Detection* (PD) module; 2) the *Points-correspondences Identification* (PcI) module; and  
956 3) the *Perspective n Point* (PnP) module. These modules are executed in a sequence at  
957 each frame captured by the camera. An estimate of the position/pose of the dot's pattern  
958 (and, therefore, of the head) is computed, only when the PD module is able to extract the  
959 positions of all the dots within the acquired image. In cases in which such condition is not  
960 satisfied, the PD algorithm signals and records the inability to successfully recover the  
961 position/pose of the pattern for the current frame.

962 The PD module is a fast feature-detection algorithm that eliminates false positives  
963 blobs and restricts the search space of point configurations by certain geometric  
964 constraints. Following the correct extraction of the dots' positions by the PD algorithm,  
965 such positions must be univocally matched to the known geometry of the pattern by the  
966 PcI algorithm. That is, all six points, as they appear in the image, must be associated to  
967 the known 3D coordinates of the corresponding dots on the 3D pattern. Finally, the  
968 position and pose of the *pattern reference system* ( $x'$ ,  $y'$ ,  $z'$ ; purple arrows in Figure 2C)  
969 with respect to the *camera reference system* ( $x$ ,  $y$ ,  $z$ ; black arrows in Figure 2C) is  
970 obtained by solving the PnP problem. The whole algorithm was designed to process the  
971 images captured by the camera in real-time, so as to output the estimated position and  
972 pose of the pattern without the need of storing the images for off-line processing. This  
973 required designing the three different modules, in such a way to maximize both the speed  
974 of processing and the accuracy of the estimates.

### 975 **Point detection (PD) module**

976 To solve the final PnP problem and estimate the pose of the dots' pattern in the camera  
977 reference system, all the possible point candidates that represent a 2D projection  
978 configuration of the pattern must be considered. This requires extracting first the  
979 positions of all six points in each frame captured by the camera. The PD module of our

980 algorithm takes care of this step by applying a Difference of Gaussians (DoG) filter,  
981 which is particularly efficient to compute, is rotationally invariant, and shows a good  
982 stability under projective transformation and illumination changes (Lowe, 1991, 2004).

983 The DoG filter is an approximation of the well-known Laplacian of Gaussian  
984 (LoG) filter. It is defined as the difference between the images resulting from filtering a  
985 given input image  $I$  with two Gaussians having different sigma,  $\sigma_1$  and  $\sigma_2$  (Jähne, 2005),  
986 i.e.:  $\text{DoG} = G(I, \sigma_1) - G(I, \sigma_2)$ , where  $G(I, \sigma)$  is the convolution of  $I$  with a Gaussian  
987 filter  $G$  with parameter  $\sigma$ . When the size of the DoG kernel matches the size of a blob-  
988 like structure in the image, the response of the filter becomes maximal. The DoG kernel  
989 can therefore be interpreted as a matching filter (Duda et al., 2001). In our  
990 implementation, the ratio  $R = \sigma_1/\sigma_2$  has been fixed to 2. Therefore,  $\sigma_1$  and  $\sigma_2$  can be  
991 written as:

$$992 \quad \sigma_1 = \sigma\sqrt{2}$$

$$993 \quad \sigma_2 = \sigma / \sqrt{2}$$

994 where  $\sigma$  can be interpreted as the size of the DoG kernel. In principle,  $\sigma$  should be set  
995 around  $r/\sqrt{2}$ , where  $r$  is the radius of a black dot as it appears on a frame imaged by the  
996 camera (Lindeberg, 1998). However, the method is quite tolerant to the variations of the  
997 distance of the dot's pattern from the camera, and the dots are correctly detected even  
998 when they are at a working distance that is half than that originally set (i.e., when their  
999 size is twice as large as  $r$ ). In addition, the software implementation of our head tracker  
1000 includes a GUI that allows manually adjusting the value of  $\sigma$ , as well as of other key  
1001 parameters (e.g.,  $I_c$  and  $score$ ; see next paragraph/sections), depending on the stability of  
1002 the tracking procedure (as visually assessed by the user, in real-time, through the GUI).

1003 After detecting the candidate dots in the image plane using the DoG filter, we  
1004 applied a *non-maxima suppression* algorithm (Canny, 1987) that rejects all candidate  
1005 locations that are not local maxima and are smaller than a contrast threshold  $I_c$ . Still,  
1006 depending on the value of  $I_c$ , a large number of false positives can be found along the  
1007 edges of the pattern. In fact, a common drawback of the DoG and LoG representations is  
1008 that local maxima can also be detected in the neighborhood of contours or straight edges,



1009 where the signal change is only in one direction. These maxima, however, are less stable,  
1010 because their localization is more sensitive to noise or small changes in neighboring  
1011 texture. A way to solve the problem of these false detections would be to analyze  
1012 simultaneously the trace and the determinant of the Hessian matrix over a neighborhood  
1013 of pixels in the image (Mikolajczyk and Schmid, 2002). The trace of the Hessian matrix  
1014 is equal to the LoG but considering simultaneously the maxima of the determinant  
1015 penalizes points for which the second derivatives detect signal-changes in only one  
1016 direction. Since a similar idea is explored in the *Harris cornerness* operator (Harris and  
1017 Stephens, 1988), our algorithm exploits the already calculated Gaussian smoothed image  
1018 and uses the efficient implementation of Harris corner detection available in the OpenCV  
1019 library (<https://opencv.org>).

1020 Given an image  $I$ , Harris cornerness operator obtains the local image structure  
1021 tensor  $H_{S_p}$  over a neighborhood of pixels  $S_p$ , where  $H_{S_p}$  is defined as:

$$H_{S_p} = \begin{bmatrix} \sum_{S_p} (\partial I / \partial x)^2 & \sum_{S_p} (\partial I / \partial x \partial I / \partial y) \\ \sum_{S_p} (\partial I / \partial x \partial I / \partial y) & \sum_{S_p} (\partial I / \partial y)^2 \end{bmatrix}$$

1022  
1023 Here,  $\partial I / \partial x$  and  $\partial I / \partial y$  are the partial derivatives of the image intensity  $I$  along the two  
1024 spatial axes  $x$  and  $y$  of the image plane, computed using a Sobel operator with an aperture  
1025 of 3 pixels – i.e., a 3x3 filter that implements a smooth, discrete approximation of a first  
1026 order derivative (Jähne, 2005; González and Woods, 2008).  $H_{S_p}$  provides a robust  
1027 distinction between edges and small blobs because the difference  $\det(H_{S_p}) - k * \text{trace}(H_{S_p})^2$   
1028 assumes values that are strictly negative on edges and positives on blob  
1029 centers. This difference depends on three parameters: 1) the aperture of the Sobel filter,  
1030 which, as mentioned above, was fixed to the minimal possible value (3) for the sake of  
1031 speed of computation; 2) the weight  $k$ , assigned to the trace of the  $H_{S_p}$  tensor, which, as  
1032 suggested in (Grauman and Leibe, 2011), was set to 0.04, i.e., the default value of the  
1033 openCV library (<https://opencv.org>); and 3) the block size of the squared neighborhood  
1034  $S_p$  that was empirically set to 9 based on some pilot tests, where it showed good stability

1035 over a broad range of working distances (note that  $S_p$  could be extrapolated from the  
1036 dots' size, the working distance between the camera and the dots' pattern and the internal  
1037 camera parameters, assuming an ideal condition of a clean white background around the  
1038 dots).

1039 To summarize, the complete PD algorithm worked as follows. First, the DoG  
1040 filter was applied to identify all candidate dots in the acquired image. Then, the non-  
1041 maxima suppression algorithm was used to prune some of the false positives around the  
1042 maxima. Finally, for each of the remaining detected dots, the Harris cornerness operator  
1043  $H_{S_p}$  was computed over the a neighborhood  $S_p$  centered on the position of each dot and,  
1044 depending on the sign of the difference  $\det(H_{S_p}) - k * \text{trace}(H_{S_p})^2$ , the dot was rejected  
1045 as a false positive or accepted as the projection on the image plane of one of the dots of  
1046 the 3D pattern. As mentioned above, the contrast threshold  $I_c$  of the non-maxima  
1047 suppression algorithm was adjustable by the user through the same GUI used to adjust  $\sigma$   
1048 of the DoG.

1049 To conclude, it should be noted that, in spite of the pruning of false detections  
1050 performed by the PD algorithm, still many spurious dots are identified in the image, in  
1051 addition to those actually present on the pattern. This is because, in general, many dot-  
1052 like features are present in any image. To further refine the identification of the actual  
1053 dots of the pattern, it is necessary to take into account their relationship, given the  
1054 geometry of the pattern itself. This is achieved by the PcI algorithm that is described in  
1055 the next section.

### 1056 **Point-correspondences identification (PcI) module**

1057 Since projective transformations maintain straight lines, aligned triplets of dots (p1, p2,  
1058 p3) in the 3D pattern must still be aligned in the images captured by the camera. Our PcI  
1059 algorithm searches all aligned triplets of dots in an image and, in order to reduce the  
1060 number of possible triplets, only those having length (i.e., distance between the external  
1061 points) smaller than  $D$  are considered. To identify a triplet, for any given pair of detected  
1062 dots, we looked whether a third, central dot was present in the proximity of the middle

1063 position in-between the pair (note that, in case of a triplet, the offset of the projected  
1064 central point with respect to the middle position is negligible, because usually the black  
1065 dots are at a much shorter distance than the working distance and the orthographic  
1066 projection can be adopted).  $D$  is automatically computed from the physical distances of  
1067 the external points of the triplets on the dot's pattern, knowing the intrinsic parameters of  
1068 the camera. It must be set to be slightly bigger (10% bigger showed to be widely  
1069 sufficient) than the maximum distance between the external points in the triplets on the  
1070 image plane, when the pattern is positioned at the maximal working distance from the  
1071 camera. This condition is achieved when the z-axis of the pattern is exactly aligned to the  
1072 optical axis of the camera.

1073         Once all candidate triplets have been identified, the algorithm looks for those  
1074 having a common external point, which corresponds to the corner of the L-shaped  
1075 arrangement of 5 coplanar dots in the pattern (Figure 2A). The search of this L's corner is  
1076 performed by considering 5-tuples of points configurations to obtain the correspondence  
1077 final assignment. In addition to collinearity, another important projective invariant is the  
1078 angular ordering (on planes facing the view direction of the imaging device). That is, if  
1079 we take three points defining a triangle, once we have established an ordering to them  
1080 (either clockwise or anti-clockwise), such ordering is maintained under any projective  
1081 transformations that looks down to the same side of the plane (Bergamasco et al., 2011).  
1082 In our framework, this implies evaluating the external product of the two vectors that  
1083 start from the common point (i.e., the L's corner) and end on the respective external  
1084 points. This establishes the orientation order and, consequently, assigns uniquely the 5  
1085 black dots correspondences between the 3D pattern and its image.

1086         Following this assignment, the 6th dot belonging to the pattern (i.e., the one  
1087 placed over the pillar; Figure 2A) is searched in proximity of L's corner. To this aim, the  
1088 maximal distance between the dot on the pillar and the L's corner ( $DP$ ) in the image  
1089 plane is automatically estimated (in the same way as  $D$ ) from the actual distance between  
1090 the two dots in the 3D pattern, knowing the camera's internal parameters and its maximal  
1091 working distance from the pattern. This yields a set of candidate pillar dots. Finding the  
1092 correct one requires evaluating each candidate dot in conjunction with the other 5  
1093 coplanar dots in terms of its ability to minimize the *reprojection error* computed by

1094 solving the PnP problem (see next section). It should be noticed that the correct  
1095 identification of the 6<sup>th</sup> point on the pillar is fundamental, since it is the only point out of  
1096 the plane that allows the PnP problem to be solved robustly. The reprojection error is  
1097 defined as the sum norm distance between the estimated positions of the dots in the  
1098 image and the projections of the physical dots of the 3D pattern on the image plane,  
1099 under a given assumed pose of the pattern and knowing the camera's calibration  
1100 parameters. More specifically, during the 6<sup>th</sup> point selection we defined the score of a  
1101 given dots' configuration as:

$$1102 \quad \textit{score} = 100 - S * \textit{Reprojection\_error}, \quad (3)$$

1103 where  $S$  is a proper scalar factor established experimentally (in our application, we set  $S =$   
1104  $5$ ). To understand the meaning of  $S$ , let's suppose, for instance, to have a distance error of  
1105 one pixel for each dot, thus yielding a reprojection error of 6. Without the  $S$  scaling factor  
1106 (i.e., with  $S = 1$ ), we would obtain a *score* of 94. However, the error on each dot is  
1107 typically well below one pixel (see next paragraph about the way to estimate the dots'  
1108 coordinates with sub-pixel accuracy) and the score would therefore be always close to  
1109 100. Hence, the need of introducing the factor  $S$  to rescale the score, so that it can range  
1110 between 90 and 100. As mentioned above, during the selection procedure, the 6<sup>th</sup> point on  
1111 the pillar was chosen among the candidates that maximized the score defined above. In  
1112 fact, each hypothesis about the position of the 6<sup>th</sup> point yielded a PnP transformation, for  
1113 which it was possible to compute the reprojection error. Note that, to eliminate some  
1114 possible ambiguities in the selection of the 6<sup>th</sup> point, we also exploited the a-priori  
1115 knowledge about the direction of the pillar (which must point toward the camera sensor).

1116 As expected, given how sensitive the PnP procedure is to small variations in the  
1117 estimated positions of the dots, we empirically verified that a pixel-level accuracy was  
1118 not sufficient to guarantee high precision and stability in our pose estimates. For this  
1119 reason, we estimated the positions of the centers of the dots at the sub-pixel level by 3-  
1120 point Gaussian approximation (Naidu and Fisher, 1991). This method considers the three  
1121 highest, contiguous intensity values (along either the  $x$  or  $y$  spatial axes) within a region  
1122 of the image that has been identified as one of the dots (i.e., a blob) and assumes that the  
1123 shape of the observed peak fits a Gaussian profile. This assumption is reasonable,

1124 because the sensor integration over a small area of pixels containing a blob, after the DoG  
1125 filtering, produces smooth profiles very similar to Gaussian profiles. If  $a$ ,  $b$  and  $c$  are the  
1126 intensity values observed at pixel positions  $x - 1$ ,  $x$  and  $x + 1$  with  $b$  having the highest  
1127 value, then the sub-pixel location ( $x_s$ ) of the peak is given by:

1128

$$x_s = x - \frac{1}{2} \frac{\ln(c) - \ln(a)}{\ln(a) + \ln(c) - 2\ln(b)}$$

1129

1130 where  $x$  is the  $x$ -coordinate of the center of the pixel with intensity value  $b$ . The same  
1131 approximation is applied to obtain the sub-pixel  $y$  coordinate  $y_s$  of the dot center.

### 1132 **Perspective n Point (PnP) module**

1133 In computer vision, the problem of estimating the position and orientation of an object  
1134 with respect to a perspective camera, given its intrinsic parameters obtained from  
1135 calibration and a set of world-to-image correspondences, is known as the Perspective-n-  
1136 Point camera pose problem (PnP) (Lowe, 1991b; Fiore, 2001). Given a number of 2D-3D  
1137 point correspondences  $m_i \leftrightarrow M_i$  (where  $m_i = [u \ v]'$  are the 2D coordinates of point  $i$   
1138 over the image plane, and  $M_i = [X \ Y \ Z]'$  are the 3D coordinates of point  $i$  in the physical  
1139 environment) and the matrix  $K$  with the intrinsic camera parameters (see definition  
1140 below), the PnP problem requires to find: 1) a rotation matrix  $R$  that defines the  
1141 orientation of the object (i.e., of the pattern reference system  $x'$ ,  $y'$ ,  $z'$  in our case) with  
1142 respect to the camera reference system  $x$ ,  $y$ ,  $z$  (see eq. 1 and Figure 2C); and 2) a  
1143 translation vector  $t$  that specifies the Cartesian coordinates of the center of the object (i.e.,  
1144 of the origin  $O'$  of the pattern reference system in our case) in the camera reference  
1145 system, such that:

$$1146 \quad \tilde{m}_i \cong K[R \ t] \tilde{M}_i \text{ for all } i, \quad (4)$$

1147 where  $\sim$  denotes homogeneous coordinates (Jähne, 2005) and  $\cong$  defines an equation up  
1148 to a scale factor. Specifically:

1149

$$\tilde{m}_i = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$[R \ t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}$$

1150 
$$\tilde{M}_i = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

1151 where  $f_x$  and  $f_y$  are the focal lengths and  $c_x$  and  $c_y$  are the coordinates of the principal point  
1152 of the camera lens.

1153 To solve the PnP problem, all methods have to face a trade-off between speed and  
1154 accuracy. Direct methods, such as Direct Linear Transform (DLT), find a solution to a  
1155 system of linear equations derived from (4) and are usually faster but less accurate, as  
1156 compared to iterative methods. On the other hand, iterative methods that explicitly  
1157 minimize a meaningful geometric error, such as the reprojection error, are more accurate  
1158 but slower (Garro et al., 2012). In our application, we adopted the method known as  
1159 EPnP (Lepetit et al., 2008), followed by an iterative refinement. The EPnP approach is  
1160 based on a non-iterative solution to the PnP problem and its computational complexity  
1161 grows linearly with  $n$ , where  $n$  is the number of point correspondences. The method is  
1162 applicable for all  $n \geq 4$  and properly handles both planar and non-planar configurations.  
1163 The central idea is to express the  $n$  3D points as a weighted sum of four virtual control  
1164 points. Since in our setup high precision is required, the output of the closed-form  
1165 solution given by the EPnP was used to initialize an iterative Levenberg-Marquardt  
1166 scheme (More, 1977), which finds the pose that minimizes the reprojection error, thus  
1167 improving the accuracy with negligible amount of additional time. Both the EPnP and the  
1168 Levenberg-Marquardt iterative scheme are available in the openCV library  
1169 (<https://opencv.org>) and are extremely efficient. The execution time in our HP  
1170 Workstation Z620 was in the order of some milliseconds.

1171

1172 **Camera calibration procedure**

1173 Camera calibration, or more precisely camera *resectioning*, is the process of estimating  
1174 the parameters of a pinhole camera model approximating the true camera that produced a  
1175 given image or a set of images. With the exception of the so-called self-calibration  
1176 methods, which try to estimate the parameter by exploiting only point correspondences  
1177 among images, a calibration-object having a known precise geometry is needed. In fact,  
1178 self-calibration cannot usually achieve an accuracy comparable with that obtained with a  
1179 calibration-object, because it needs to estimate a large number of parameters, resulting in  
1180 a much harder mathematical problem (Zhang, 2000).

1181 Much progress has been done, starting in the photogrammetry community, and  
1182 more recently in the field of computer vision, in terms of developing object-based  
1183 calibration methods. In general, these approaches can be classified in two major  
1184 categories, based on the number of dimensions of the calibration objects: 1) 3D object-  
1185 based calibration, where camera calibration is performed by observing a calibration  
1186 object whose geometry in 3-D space is known with very good precision; and 2) 2D plane-  
1187 based calibration, which is based on imaging a planar pattern shown at a few different  
1188 orientations. In our application, we adopted this second option, because it has proven to  
1189 be the best choice in most situations, given its ease of use and good accuracy.  
1190 Specifically, we adopted the method of (Zhang, 2000), available in the OpenCV library  
1191 (<https://opencv.org>), which, in its iterative process, also estimates some lens distortion  
1192 coefficients (see the next section for a discussion on the distortion).

1193 The fundamental equation to achieve the calibration is the same of the PnP  
1194 problem, [i.e., equation (4)], and the iterative solution is based on minimizing the  
1195 reprojection error defined in (3). However, in the case of the camera calibration, also the  
1196 matrix  $K$  with the parameters is unknown, in addition to  $R$  and  $t$ . As such, solving the  
1197 equation is, in principle, harder. However, the algorithm used to minimize the  
1198 reprojection error does not need to run in real-time, since the calibration is performed  
1199 before the camera is used for head tracking. In addition, the point correspondences  
1200  $m_i \leftrightarrow M_i$  over which the error is computed and minimized are the order of several  
1201 hundreds, which makes the estimation of the parameters very robust and reliable. In our  
1202 application, these points were the intersections of the  $9 \times 7$  squares of a planar

1203 checkerboard (shown in Figure 4-figure supplement 1A) imaged in 15 different  
 1204 poses/positions. Specifically, a few snapshots were taken centrally at different distances  
 1205 from the camera, others spanned the image plane to sample the space where the radial  
 1206 distortion is more prominent, and finally (and more importantly) other snapshots were  
 1207 taken from different angles of orientation with respect to the image plane (Zhang, 2000)

1208 To measure the effect of changing the calibration images over the pose estimation,  
 1209 we collected a set of 80 images of the calibration checkerboard at different orientations.  
 1210 50 random subsamples (without replacement), each composed of 50% of the total images,  
 1211 were used to calibrate the system, thus yielding 50 different calibrations. In Table 1, we  
 1212 report the mean and standard deviation of the internal parameters and the lens distortion  
 1213 coefficients obtained from such calibrations. In Table 2, we report the averages and  
 1214 standard deviations of the radial and tangential part contributions of the distortion. Some  
 1215 of the parameters reported in these tables (i.e.,  $f_x$ ,  $f_y$ ,  $c_x$  and  $c_y$ ) have been defined in eq.  
 1216 (4), while other parameters ( $k_1$ ,  $k_2$ ,  $k_3$ ,  $p_1$  and  $p_2$ ) are distortion coefficients described in  
 1217 the next paragraphs.

1218

	$f_x$	$f_y$	$C_x$	$c_y$	$k_1$	$k_2$	$p_1$	$p_2$	$k_3$	Rms [pixel]
<b>mean</b>	2701.7	2707.4	636.6	508.9	-0.396	2.23	0.00098	-0.0019	-26.37	0.10
<b>Std</b>	2.51	1.98	4.73	4.29	0.012	0.78	0.00018	0.00025	15.50	0.00094

1219

1220 **Table1.** Mean values and standard deviations of the internal camera parameters  
 1221 and distortion coefficients. The reprojection error, namely the distance between  
 1222 the square's corners in the images and their points reprojected after the  
 1223 calibration, is reported in the last column.

1224

	Radial dist.	Tangential Dist.
<b>Mean</b>	-0.02048	-0.00020
<b>Std</b>	0.0010	2.49e-05



1225

1226 **Table 2.** Mean values and standard deviations of the two components of the  
1227 distortion, i.e. the radial and the tangential part. The radial distortion of -0.02048  
1228 indicates that the pixels in the corners of the image appear about 2% closer to  
1229 the center of the image (barrel distortion).

1230

1231 The focal lengths  $f_x$  and  $f_y$  (Table 1) showed a good stability. The coordinates of the  
1232 principal point  $(c_x, c_y)$ , as it is well known, are some of the hardest parameters to  
1233 estimate and, in our case, had a standard deviation of approximately 4.5 pixels. In any  
1234 event, they are additive terms and, as such, do not affect distance and angle  
1235 measurements. The remaining parameters describe the distortion introduced by the lens  
1236 during the image formation. We quantified the effect of distortion, both the radial and the  
1237 tangential part, on the pose estimation by considering our specific setup.

1238 From eq. (4), after the roto-translation transformation  $[x \ y \ z] = [R \ t][X \ Y \ Z \ 1]$ ,  
1239 we have the 2d homogeneous coordinates

$$1240 \quad x' = x/z, y' = y/z.$$

1241 Then, the distorted coordinates  $x''$  and  $y''$  can be modeled by the parameters  $k_1, k_2, k_3, p_1$   
1242 and  $p_2$  using the following equation:

$$x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x' y' + p_2 (r^2 + 2x'^2)$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 x' y' + p_1 (r^2 + 2y'^2)$$

1243 where  $r^2 = x'^2 + y'^2$ . The final 2D image coordinates are then obtained by:

$$s[u \ v \ 1] = K[x'' \ y'' \ 1]$$

1244 In the worst case scenario, namely in the corners of the image where the distortion is at  
1245 its maximum,  $x'$  and  $y'$  can be approximated by  $dS_x/2f$  and  $dS_y/2f$ , where  $f$  is the focal  
1246 length of the lens and  $dS_x$  and  $dS_y$  are the x and y dimensions of the CCD sensor. By  
1247 simple geometrical considerations, since our camera mounts a 1/3" CMOS sensor and the  
1248 lens has a focal length of 12.5mm,  $r^2$  results equal to 0.0576. The distortion corrections  
1249 were then calculated for every sample considering this worst case. The radial distortion

1250 correction  $(k_1r^2 + k_2r^4 + k_3r^6)$  dominates the tangential part  $(2p_1x'y' + p_2(r^2 +$   
1251  $2x'^2))$  but, more importantly, it has a standard deviation of just about 0.1% (see Table 2).

1252 In this framework, it is clear that the calibration is very stable, with parameters  
1253 fluctuating around 0,1% for  $fx, fy$  and the distortion coefficients, whereas  $cx$  and  $cy$  do not  
1254 affect the displacement or angular measurements. Vice versa, an accurate measurement  
1255 of the physical distances between the dots of pattern is crucial. We verified that an error  
1256 of 5% in the dot distances produces errors of 5% in the pose estimates (for example, if the  
1257 distances are estimated 5% bigger than the truth, the dots' pattern is estimated at a pose  
1258 5% more distant to the camera) and, consequently, the translation measurements are  
1259 affected.

1260

1261 **Acknowledgments.** This work was supported by a Human Frontier Science Program  
1262 Grant to DZ (contract n. RGP0015/2013), a European Research Council Consolidator  
1263 Grant to DZ (project n. 616803-LEARN2SEE) and a FSE POR Regione autonoma FVG  
1264 Program Grant (HEaD - Higher education and development) to WV. We thank Marco  
1265 Gigante for his technical support in designing and building the custom assemblies used to  
1266 validate the head tracker.

1267

1268 **Competing interests.** None of the authors has financial and non-financial competing  
1269 interests of any nature.

1270

1271

## 1272 **References**

1273 Acharya L, Aghajan ZM, Vuong C, Moore JJ, Mehta MR (2016) Causal Influence of  
1274 Visual Cues on Hippocampal Directional Selectivity. *Cell* 164:197–207.

1275 Alemi-Neissi A, Rosselli FB, Zoccolan D (2013) Multifetural Shape Processing in Rats  
1276 Engaged in Invariant Visual Object Recognition. *J Neurosci* 33:5939–5956.

1277 Aragão R da S, Rodrigues MAB, de Barros KMFT, Silva SRF, Toscano AE, de Souza  
1278 RE, Manhães-de-Castro R (2011) Automatic system for analysis of locomotor  
1279 activity in rodents—A reproducibility study. *J Neurosci Methods* 195:216–221.

- 1280 Attanasi A, Cavagna A, Castello LD, Giardina I, Grigera TS, Jelić A, Melillo S, Parisi L,  
1281 Pohl O, Shen E, Viale M (2014) Information transfer and behavioural inertia in  
1282 starling flocks. *Nat Phys* 10:691–696.
- 1283 Bassett JP, Taube JS (2001) Neural Correlates for Angular Head Velocity in the Rat  
1284 Dorsal Tegmental Nucleus. *J Neurosci* 21:5740–5751.
- 1285 Berg HC, Brown DA (1972) Chemotaxis in *Escherichia coli* analysed by Three-  
1286 dimensional Tracking. *Nature* 239:500.
- 1287 Bergamasco F, Albarelli A, Torsello A (2011) Image-Space Marker Detection and  
1288 Recognition Using Projective Invariants In Visualization and Transmission 2011  
1289 International Conference on 3D Imaging, Modeling, Processing , p. 381–388.
- 1290 Bossens C, Beeck O de, P H (2016) Linear and Non-Linear Visual Feature Learning in  
1291 Rat and Humans. *Front Behav Neurosci* 10.
- 1292 Canny J (1987) A Computational Approach to Edge Detection In M. A. Fischler & O.  
1293 Firschein, eds. *Readings in Computer Vision* San Francisco (CA): Morgan  
1294 Kaufmann, p. 184–203. Available at:  
1295 <http://www.sciencedirect.com/science/article/pii/B9780080515816500246>  
1296 [Accessed March 27, 2019].
- 1297 Cavagna A, Conti D, Creato C, Castello LD, Giardina I, Grigera TS, Melillo S, Parisi L,  
1298 Viale M (2017) Dynamic scaling in natural swarms. *Nat Phys* 13:914–918.
- 1299 De Keyser R, Bossens C, Kubilius J, Beeck HPO de (2015) Cue-invariant shape  
1300 recognition in rats as tested with second-order contours. *J Vis* 15:14–14.
- 1301 Djurdjevic V, Ansuini A, Bertolini D, Macke JH, Zoccolan D (2018) Accuracy of Rats in  
1302 Discriminating Visual Objects Is Explained by the Complexity of Their  
1303 Perceptual Strategy. *Curr Biol* 28:1005-1015.e5.
- 1304 Duda RO, Hart PE, Stork DG (2001) *Pattern classification*. Wiley.
- 1305 Finkelstein A, Derdikman D, Rubin A, Foerster JN, Las L, Ulanovsky N (2015) Three-  
1306 dimensional head-direction coding in the bat brain. *Nature* 517:159–164.
- 1307 Fiore PD (2001) Efficient linear solution of exterior orientation. *IEEE Trans Pattern Anal*  
1308 *Mach Intell* 23:140–148.
- 1309 Fyhn M, Molden S, Witter MP, Moser EI, Moser M-B (2004) Spatial Representation in  
1310 the Entorhinal Cortex. *Science* 305:1258–1264.
- 1311 Garcia-Perez E, Mazzone A, Zoccolan D, Robinson HP, Torre V (2005) Statistics of  
1312 decision making in the leech. *J Neurosci* 25:2597–608.

- 1313 Garro V, Crosilla F, Fusiello A (2012) Solving the PnP Problem with Anisotropic  
1314 Orthogonal Procrustes Analysis In Visualization Transmission 2012 Second  
1315 International Conference on 3D Imaging, Modeling, Processing , p. 262–269.
- 1316 González RC, Woods RE (2008) Digital Image Processing. Pearson/Prentice Hall.
- 1317 Gower PSDJC, Gower JC, Dijksterhuis GB, Dijksterhuis C and MIA and FIBVWU and  
1318 RC and D of M and MRF of EGB (2004) Procrustes Problems. OUP Oxford.
- 1319 Grauman K, Leibe B (2011) Visual Object Recognition 1 edition. San Rafael, Calif.:  
1320 Morgan & Claypool Publishers.
- 1321 Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005) Microstructure of a spatial  
1322 map in the entorhinal cortex. Nature 436:801.
- 1323 Harris C, Stephens M (1988) A Combined Corner and Edge Detector. In C. J. Taylor, ed.  
1324 Proceedings of the Alvey Vision Conference Alvey Vision Club, p. 23.1-23.6.
- 1325 Horner AE, Heath CJ, Hvoslef-Eide M, Kent BA, Kim CH, Nilsson SRO, Alsiö J,  
1326 Oomen CA, Holmes A, Saksida LM, Bussey TJ (2013) The touchscreen operant  
1327 platform for testing learning and memory in rats and mice. Nat Protoc 8:1961–  
1328 1984.
- 1329 Jähne B (2005) Digital Image Processing. Springer Science & Business Media.
- 1330 Kaliukhovich DA, Op de Beeck H (2018) Hierarchical stimulus processing in rodent  
1331 primary and lateral visual cortex as assessed through neuronal selectivity and  
1332 repetition suppression. J Neurophysiol 120:926–941.
- 1333 Kimmel DL, Mammo D, Newsome WT (2012) Tracking the eye non-invasively:  
1334 simultaneous comparison of the scleral search coil and optical tracking techniques  
1335 in the macaque monkey. Front Behav Neurosci 6 Available at:  
1336 <https://www.frontiersin.org/articles/10.3389/fnbeh.2012.00049/full> [Accessed  
1337 March 29, 2019].
- 1338 Knutsen PM, Derdikman D, Ahissar E (2005) Tracking Whisker and Head Movements in  
1339 Unrestrained Behaving Rodents. J Neurophysiol 93:2294–2301.
- 1340 Kropff E, Carmichael JE, Moser M-B, Moser EI (2015) Speed cells in the medial  
1341 entorhinal cortex. Nature 523:419–424.
- 1342 Kurnikova A, Moore JD, Liao S-M, Deschênes M, Kleinfeld D (2017) Coordination of  
1343 Orofacial Motor Actions into Exploratory Behavior by Rat. Curr Biol 27:688–  
1344 696.
- 1345 Kurylo DD, Chung C, Yeturo S, Lanza J, Gorskaya A, Bukhari F (2015) Effects of  
1346 contrast, spatial frequency, and stimulus duration on reaction time in rats. Vision  
1347 Res 106:20–26.

- 1348 Kurylo DD, Yeturo S, Lanza J, Bukhari F (2017) Lateral masking effects on contrast  
1349 sensitivity in rats. *Behav Brain Res* 335:1–7.
- 1350 Lepetit V, Moreno-Noguer F, Fua P (2008) EPnP: An Accurate O(n) Solution to the PnP  
1351 Problem. *Int J Comput Vis* 81:155.
- 1352 Lindeberg T (1998) Feature Detection with Automatic Scale Selection. *Int J Comput Vis*  
1353 30:79–116.
- 1354 Lowe D (1991) Fitting Parameterized Three-Dimensional Models to Images. *IEEE Trans*  
1355 *Pattern Anal Mach Intell* 13:441–450.
- 1356 Lowe DG (2004) Distinctive Image Features from Scale-Invariant Keypoints. *Int J*  
1357 *Comput Vis* 60:91–110.
- 1358 Mar AC, Horner AE, Holmes A, Kent BA, Kim CH, Alsiö J, Saksida LM, Nilsson SRO,  
1359 Bussey TJ (2013) The touchscreen operant platform for assessing executive  
1360 function in rats and mice. *Nat Protoc* 8:1985.
- 1361 Matteucci G, Marotti RB, Riggi M, Rosselli FB, Zoccolan D (2019) Nonlinear  
1362 Processing of Shape Information in Rat Lateral Extrastriate Cortex. *J Neurosci*  
1363 39:1649–1670.
- 1364 Mazzoni A, Garcia-Perez E, Zoccolan D, Graziosi S, Torre V (2005) Quantitative  
1365 characterization and classification of leech behavior. *J Neurophysiol* 93:580–93.
- 1366 Mersch DP, Crespi A, Keller L (2013) Tracking Individuals Shows Spatial Fidelity Is a  
1367 Key Regulator of Ant Social Organization. *Science* 340:1090–1093.
- 1368 Mikolajczyk K, Schmid C (2002) An Affine Invariant Interest Point Detector In A.  
1369 Heyden, G. Sparr, M. Nielsen, & P. Johansen, eds. *Computer Vision — ECCV*  
1370 2002 *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, p. 128–  
1371 142.
- 1372 More J (1977) The Levenberg-Marquardt algorithm: Implementation and theory In G.  
1373 Watson, ed. *Lecture Notes in Mathematics* 630: *Numerical Analysis* Springer  
1374 Verlag, p. 105–116.
- 1375 Moser EI, Kropff E, Moser M-B (2008) Place Cells, Grid Cells, and the Brain’s Spatial  
1376 Representation System. *Annu Rev Neurosci* 31:69–89.
- 1377 Moser M-B, Rowland DC, Moser EI (2015) Place Cells, Grid Cells, and Memory. *Cold*  
1378 *Spring Harb Perspect Biol* 7:a021808.
- 1379 Naidu DK, Fisher RB (1991) A Comparative Analysis of Algorithms for Determining the  
1380 Peak Position of a Stripe to Sub-pixel Accuracy In P. Mowforth, ed. *BMVC91*  
1381 Springer London, p. 217–225.

- 1382 Nikbakht N, Tafreshiha A, Zoccolan D, Diamond ME (2018) Supralinear and  
1383 Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and  
1384 Neuronal Mechanisms. *Neuron* 97:626-639.e8.
- 1385 O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence  
1386 from unit activity in the freely-moving rat. *Brain Res* 34:171–175.
- 1387 O'Keefe J, Nadel L (1978) *The Hippocampus as a Cognitive Map*. Oxford: New York:  
1388 Oxford University Press.
- 1389 Pasquet MO, Tihy M, Gurgeon A, Pompili MN, Godsil BP, Léna C, Dugué GP (2016)  
1390 Wireless inertial measurement of head kinematics in freely-moving rats. *Sci Rep*  
1391 6:35689.
- 1392 Payne HL, Raymond JL (2017) Magnetic eye tracking in mice. *eLife* Available at:  
1393 <https://elifesciences.org/articles/29222> [Accessed August 21, 2018].
- 1394 Perkon I, Košir A, Itskov PM, Tasič J, Diamond ME (2011) Unsupervised quantification  
1395 of whisking and head movement in freely moving rodents. *J Neurophysiol*  
1396 105:1950–1962.
- 1397 Pinnell RC, Almajidy RK, Kirch RD, Cassel JC, Hofmann UG (2016) A Wireless EEG  
1398 Recording Method for Rat Use inside the Water Maze. *PLOS ONE* 11:e0147730.
- 1399 Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised Spike Detection and Sorting  
1400 with Wavelets and Superparamagnetic Clustering. *Neural Comput* 16:1661–1687.
- 1401 Remmel RS (1984) An inexpensive eye movement monitor using the scleral search coil  
1402 technique. *IEEE Trans Biomed Eng* 31:388–90.
- 1403 Rigosa J, Lucantonio A, Noselli G, Fassihi A, Zorzin E, Manzino F, Pulecchi F, Diamond  
1404 ME (2017) Dye-enhanced visualization of rat whiskers for behavioral studies.  
1405 *eLife* Available at: <https://elifesciences.org/articles/25290> [Accessed August 21,  
1406 2018].
- 1407 Rosselli FB, Alemi A, Ansuini A, Zoccolan D (2015) Object similarity affects the  
1408 perceptual strategy underlying invariant visual object recognition in rats. *Front*  
1409 *Neural Circuits* 9:10.
- 1410 Sargolini F, Fyhn M, Hafting T, McNaughton BL, Witter MP, Moser M-B, Moser EI  
1411 (2006) Conjunctive Representation of Position, Direction, and Velocity in  
1412 Entorhinal Cortex. *Science* 312:758–762.
- 1413 Sawinski J, Wallace DJ, Greenberg DS, Grossmann S, Denk W, Kerr JND (2009)  
1414 Visually evoked activity in cortical cells imaged in freely moving animals. *Proc*  
1415 *Natl Acad Sci* 106:19557–19562.

- 1416 Stackman RW, Taube JS (1998) Firing Properties of Rat Lateral Mammillary Single  
1417 Units: Head Direction, Head Pitch, and Angular Head Velocity. *J Neurosci*  
1418 18:9020–9037.
- 1419 Stahl JS, van Alphen AM, De Zeeuw CI (2000) A comparison of video and magnetic  
1420 search coil recordings of mouse eye movements. *J Neurosci Methods* 99:101–10.
- 1421 Stirman JN, Townsend LB, Smith SL (2016) A touchscreen based global motion  
1422 perception task for mice. *Vision Res* 127:74–83.
- 1423 Szuts TA, Fadeyev V, Kachiguine S, Sher A, Grivich MV, Agrochão M, Hottowy P,  
1424 Dabrowski W, Lubenov EV, Siapas AG, Uchida N, Litke AM, Meister M (2011)  
1425 A wireless multi-channel neural amplifier for freely moving animals. *Nat*  
1426 *Neurosci* 14:263–269.
- 1427 Tafazoli S, Di Filippo A, Zoccolan D (2012) Transformation-Tolerant Object  
1428 Recognition in Rats Revealed by Visual Priming. *J Neurosci* 32:21–34.
- 1429 Tafazoli S, Safaai H, Franceschi GD, Rosselli FB, Vanzella W, Riggi M, Buffolo F,  
1430 Panzeri S, Zoccolan D (2017) Emergence of transformation-tolerant  
1431 representations of visual objects in rat lateral extrastriate cortex. *eLife* 6:e22794.
- 1432 Taube JS (2007) The Head Direction Signal: Origins and Sensory-Motor Integration.  
1433 *Annu Rev Neurosci* 30:181–207.
- 1434 Tort ABL, Neto WP, Amaral OB, Kazlauskas V, Souza DO, Lara DR (2006) A simple  
1435 webcam-based approach for the measurement of rodent locomotion and other  
1436 behavioural parameters. *J Neurosci Methods* 157:91–97.
- 1437 Tsoar A, Nathan R, Bartan Y, Vyssotski A, Dell’Omo G, Ulanovsky N (2011) Large-  
1438 scale navigational map in a mammal. *Proc Natl Acad Sci* 108:E718–E724.
- 1439 Vermaercke B, Op de Beeck HP (2012) A Multivariate Approach Reveals the Behavioral  
1440 Templates Underlying Visual Discrimination in Rats. *Curr Biol* 22:50–55.
- 1441 Vinken K, Van den Bergh G, Vermaercke B, Beeck O de, P H (2016) Neural  
1442 Representations of Natural and Scrambled Movies Progressively Change from  
1443 Rat Striate to Temporal Cortex. *Cereb Cortex* 26:3310–3322.
- 1444 Vinken K, Vermaercke B, Op de Beeck HP (2014) Visual Categorization of Natural  
1445 Movies by Rats. *J Neurosci* 34:10645–10658.
- 1446 Vinken K, Vogels R, Op de Beeck H (2017) Recent Visual Experience Shapes Visual  
1447 Processing in Rats through Stimulus-Specific Adaptation and Response  
1448 Enhancement. *Curr Biol* 27:914–919.

- 1449 Wallace DJ, Greenberg DS, Sawinski J, Rulla S, Notaro G, Kerr JND (2013) Rats  
1450 maintain an overhead binocular field at the expense of constant fusion. *Nature*  
1451 498:65–69.
- 1452 Yartsev MM, Ulanovsky N (2013) Representation of Three-Dimensional Space in the  
1453 Hippocampus of Flying Bats. *Science* 340:367–372.
- 1454 Yu Y, Hira R, Stirman JN, Yu W, Smith IT, Smith SL (2018) Mice use robust and  
1455 common strategies to discriminate natural scenes. *Sci Rep* 8:1379.
- 1456 Zhang Z (2000) A Flexible New Technique for Camera Calibration. *IEEE Trans Pattern*  
1457 *Anal Mach Intell* 22 Available at: [https://www.microsoft.com/en-](https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/)  
1458 [us/research/publication/a-flexible-new-technique-for-camera-calibration/](https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/)  
1459 [Accessed December 14, 2018].
- 1460 Zoccolan D (2015) Invariant visual object recognition and shape processing in rats.  
1461 *Behav Brain Res* 285:10–33.
- 1462 Zoccolan D, Di Filippo A (2018) Chapter 5 - Methodological Approaches to the  
1463 Behavioural Investigation of Visual Perception in Rodents In A. Ennaceur & M.  
1464 A. de Souza Silva, eds. *Handbook of Behavioral Neuroscience Handbook of*  
1465 *Object Novelty Recognition*. Elsevier, p. 69–101.
- 1466 Zoccolan D, Graham BJ, Cox DD (2010) A self-calibrating, camera-based eye tracker for  
1467 the recording of rodent eye movements. *Front Neurosci* 4:193.
- 1468 Zoccolan D, Oertelt N, DiCarlo JJ, Cox DD (2009) A rodent model for the study of  
1469 invariant visual object recognition. *Proc Natl Acad Sci U A* 106:8748–53.
- 1470