



**HAL**  
open science

# Task Engagement Enhances Population Encoding of Stimulus Meaning in Primary Auditory Cortex

Sophie Bagur, Martin Averseng, Diego Elgueda, Stephen David, Jonathan Fritz, Pingbo Yin, Shihab Shamma, Yves Boubenec, Srdjan Ostojic

## ► To cite this version:

Sophie Bagur, Martin Averseng, Diego Elgueda, Stephen David, Jonathan Fritz, et al.. Task Engagement Enhances Population Encoding of Stimulus Meaning in Primary Auditory Cortex. *Nature Communications*, 2017, 9 (1), pp.2529. <10.1101/240556>. <hal-03923652>

**HAL Id: hal-03923652**

**<https://hal.science/hal-03923652v1>**

Submitted on 4 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Task Engagement Enhances Population Encoding of Stimulus Meaning in Primary Auditory Cortex

Sophie Bagur<sup>1</sup>, Martin Averseng<sup>2</sup>, Diego Elgueda<sup>3</sup>, Stephen David<sup>4</sup>, Jonathan Fritz<sup>3</sup>, Pingbo Yin<sup>3</sup>, Shihab Shamma<sup>2,3</sup>, Yves Boubenec<sup>2\*</sup>, Srdjan Ostojic<sup>5\*</sup>

<sup>1</sup> Brain Plasticity Unit, Équipe MOBS, CNRS UMR 8249, École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, Paris, France.

<sup>2</sup> Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL Research University, CNRS, 75005 Paris, France

<sup>3</sup> Neural Systems Laboratory, Institute for Systems Research & Electrical and Computer Engineering, University of Maryland in College Park, MD, USA.

<sup>4</sup> Laboratory of Brain, Hearing and Behavior, Oregon Health & Science University, OR, USA.

<sup>5</sup> Group for Neural Theory, Laboratoire de Neurosciences Cognitives, INSERM U960, École Normale Supérieure, PSL Research University, CNRS, 75005 Paris, France.

\* Equal contribution

## Abstract

The main functions of primary sensory cortical areas are classically considered to be the extraction and representation of stimulus features. In contrast, higher cortical sensory association areas are thought to be responsible for combining these sensory representations with internal motivations and learnt associations. These regions generate appropriate neural responses that are maintained until a motor command is executed. Within this framework, responses of the primary sensory areas during task performance are expected to carry less information about the behavioral meaning of the stimulus than higher sensory, association, motor and frontal cortices. Here we demonstrate instead that the neuronal population responses in the early primary auditory cortex (A1) display many aspects of responses generally associated with higher-level areas. A1 activity was recorded in awake ferrets while they were either passively listening or actively discriminating two periodic click trains of different rates in a Go/No-Go paradigm. By applying population-level dimensionality reduction techniques, we found that task-engagement induced a shift in the nature of the encoding from a sensory-driven representation of the two stimuli to a behaviorally relevant representation of the two categories that specifically enhances the target stimulus. We demonstrate that this shift in encoding relies partly on a novel mechanism of change in spontaneous activity patterns upon engagement in the task. We show that this population-level representation of stimuli in A1 population activity bears strong similarities to responses in the frontal cortex, but appears earlier following stimulus presentation. Analysis of neural activity recorded in various Go/No-Go tasks, with different sounds and reinforcement paradigms, reveals that this striking population-level enhancement of target representation is a general property of task engagement. These findings indicate that primary sensory cortices play a highly flexible role in the processing of incoming stimuli and implement a crucial change in the structure of population activity in order to extract task-relevant information during behavior.

## 51 Introduction

52

53 How and where in the brain are sensory representations transformed into abstract  
54 percepts? Classical anatomical and physiological studies have suggested that this  
55 transformation occurs progressively along a cortical hierarchy. Primary sensory areas  
56 are commonly believed to process and extract high-level physical properties of  
57 stimuli, such as orientations of visual bars in the primary visual cortex or abstract  
58 sound features in the primary auditory cortex<sup>1,2</sup>. These fundamental sensory features  
59 are then integrated and interpreted as behaviorally meaningful sensory objects in  
60 sensory scenes, and relayed to higher cortical areas, which extract increasingly task-  
61 relevant abstract information. Prefrontal, parietal and premotor areas lie at the apex  
62 of the hierarchy<sup>3,4</sup>. They integrate inputs from different sensory modalities, transform  
63 sensory information into categorical percepts and decisions, and store them in  
64 working memory until the time when the appropriate motor action needs to be  
65 executed<sup>5,6</sup>.

66

67 According to this classical feedforward picture, primary sensory areas are often  
68 considered as playing a largely static role in extracting and encoding high-level  
69 stimulus physical attributes<sup>7-10</sup>. However a number of recent studies in awake,  
70 behaving animals have challenged this view, and shown that the information  
71 represented in primary areas in fact strongly depends on the behavioral state of the  
72 animal. Motor activity, arousal, learning and task-engagement have been found to  
73 strongly modulate responses in primary visual, somatosensory, and auditory cortices  
74<sup>11-25</sup>. Effects of task-engagement have been particularly investigated in the auditory  
75 cortex, where it was found that receptive fields of primary auditory cortex neurons  
76 adapt rapidly to behavioral demands when animals engage in various types of  
77 auditory discrimination tasks<sup>26-30</sup>. These observations have been interpreted as  
78 signatures of highly flexible sensory representations in primary cortical areas, and  
79 they raise the possibility that these areas may be performing computations more  
80 complex than simple extraction and transmission of processed stimulus features to  
81 higher-order regions.

82

83 An important limitation of many previous studies<sup>26-30</sup> is that they relied mostly on  
84 single-cell analyses, which characterized the selectivity of individual neurons to  
85 sensory stimuli. Here we show that simple population analyses reveal that task-  
86 engagement induces a shift in the primary auditory cortex from a sensory-driven  
87 representation to a representation of the behavioral meaning of stimuli, analogous to  
88 the one found in the frontal cortex. We first analyzed the responses during a temporal  
89 auditory discrimination task, in which ferrets had to distinguish between Go  
90 (Reference) and No-Go (Target) stimuli corresponding to click trains of different  
91 rates. The activity of the same neural population was recorded when the animals  
92 were engaged in the task, and when they passively listened to the same stimuli. Both  
93 single cell and population analyses showed that task-engagement decreased the  
94 accuracy of encoding the physical attributes of stimuli. Population, but not single-cell,  
95 analyses however revealed that task-engagement induced a shift towards an  
96 asymmetric representation of the two stimuli that enhanced target-evoked activity in  
97 the subspace of optimal decoding. This shift was in part enabled by a novel  
98 mechanism based on the change in the pattern of spontaneous activity during task  
99 engagement.

100

101 Performing identical analyses developed on this task to independent data sets  
102 collected in A1 during other behavioral discrimination tasks demonstrated that these  
103 findings can be well generalized, independently of the type of stimuli, behavioral  
104 paradigm or reward contingencies. Specifically, in all tasks, we found an enhanced  
105 representation of the target stimuli, defined as those stimuli that induced a change in  
106 the animal's ongoing behavior. Furthermore, in tasks that displayed a shift in the  
107 spontaneous firing rates of neurons, this task-adaptive encoding was partly mediated  
108 by a re-patterning of the population spontaneous activity, offering a functional  
109 interpretation for this previously observed phenomena of task-evoked changes in  
110 spontaneous activity<sup>19</sup>.

111  
112 Finally, a comparison between population activity in A1 and single-cell recordings in  
113 the frontal cortex revealed strong similarities. However, the target-driven  
114 representation of behavioral meaning appeared in A1 very rapidly following stimulus  
115 presentation, hence it was unlikely to be solely due to immediate top-down influences  
116 from frontal cortex. Altogether, our results suggest that task-relevant, abstracted  
117 information is present in primary sensory cortices, and can be read out by neurons in  
118 higher order cortices.

119  
120

## 121 RESULTS

122  
123

### 124 Task engagement degrades the encoding of stimulus physical features in A1

125  
126

127 We recorded the activity of 370 units in the primary auditory cortex (A1) of two awake  
128 ferrets in response to periodic click trains. The animals were trained using a  
129 conditioned avoidance paradigm<sup>26</sup> to lick water from a spout during the presentation  
130 of a class of reference stimuli and to stop licking following a target stimulus (Animal 1:  
131 83% hit +/- 3% s.e.m; Animal 2: 69% hit +/- 5% s.e.m) (Fig. 1a; see Methods). Target  
132 stimuli thus required a change in the ongoing behavioral output while reference  
133 stimuli did not. Each animal was trained to discriminate low vs high click rates, but  
134 the precise rates of reference and target click trains changed in every session. The  
135 category choice was opposite in the two animals to avoid confounding effects of  
136 stimulus rates (low/high) and behavioral category (reference/target). Thus, the target  
137 for one ferret was high click train rates, and the target for the other ferret was low  
138 click train rates. In each session, the activity of the same set of single units was  
139 recorded during active behavior (task-engaged condition) and during passive  
140 presentations of the same set of auditory stimuli before and after behavior (passive  
141 conditions).

142  
143

144 We first examined how auditory cortex responses and stimulus encoding depended  
145 on the behavioral state of the animal. In agreement with previous studies<sup>14,19</sup>,  
146 spontaneous activity often increased in the task-engaged condition, while stimulus-  
147 evoked activity was often suppressed (Fig. 1b). To quantify the changes in activity  
148 over the population, we used a modulation index of mean firing-rates between  
149 passive and task-engaged conditions, estimated in different epochs (Fig. 1c; see  
150 Methods). Spontaneous activity before stimulus presentation increased in the  
engaged condition (n=370 units, P<0.0001), while baseline-corrected stimulus-  
evoked activity did not change overall (n=370 units, P=0.94). These changes in

151 average activity suggested that the signal-to-noise (SNR) ratio between stimulus-  
152 evoked and spontaneous activity paradoxically decreased when the animals  
153 engaged in the task.

154

155 To quantify in a more refined manner the timing of neural responses with respect to  
156 click-times, we computed the vector strengths of individual unit responses, a  
157 standard measure of phase-locked activity evoked by click trains<sup>12,31</sup>. Vector  
158 strengths quantify the amount of entrainment of the neural response to the clicks,  
159 and range from 1 for responses fully locked to clicks to 0 for responses independent  
160 of click timing. A vast majority of neurons (Passive Ref/Targ: 80%, 81% and Active  
161 Ref/Targ: 84%, 81%) displayed statistically significant vector strengths in both  
162 conditions. However vector strength decreased in the engaged condition compared  
163 to the passive condition (Fig. 1c;  $n=574$  (287 units, 2 sounds),  $P<0.0001$ ),  
164 independently of the rate of the click train and the identity of the stimuli (Fig. S1). This  
165 reduction in stimulus-entrainment further suggested that task engagement degraded  
166 the encoding of click-times in A1.

167

168 The change in activity between passive and task-engaged conditions was  
169 heterogeneous across the neural population. While stimulus-entrainment was on  
170 average reduced in the engaged condition, a minority of neurons increased their  
171 responses. One possibility is that such changes reflect an increased sparseness of  
172 the neural code. Under this hypothesis, the stimuli are represented by smaller pools  
173 of neurons in the task-engaged condition, but in a more reliable manner. To address  
174 this possibility, we built optimal decoders that reconstructed click timings from the  
175 activity of all simultaneously recorded neurons, in a trial-by-trial manner (Fig. 1d,  
176 Methods). We found that the reconstruction accuracy decreased in the task-engaged  
177 condition compared to the passive condition (Fig. 1e-g), confirming that encoding of  
178 click-times decreased during behavior.

179

180 In summary, the fine physical features of the behaviorally relevant stimuli became  
181 less faithfully represented by A1 activity when the animals were engaged in this  
182 discrimination task.

183

184

185 **During sound presentation target and reference stimuli can be equally**  
186 **classified from A1 responses in passive and engaged conditions**

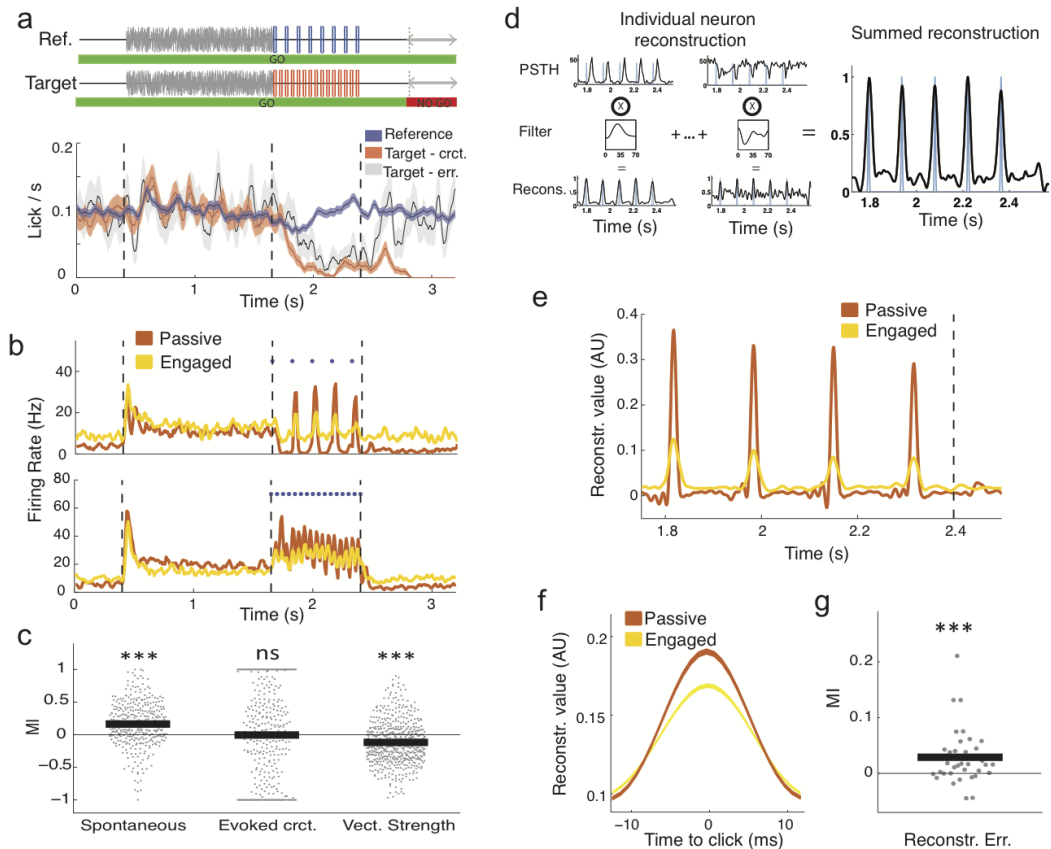
187

188 In the task-engaged condition, the animals were required to determine whether the  
189 rate of each presented click train was high or low. They needed to make a categorical  
190 decision about the stimuli and correctly associate them with the required actions,  
191 before using that information to drive behavior. We therefore asked to what extent the  
192 two classes of stimuli could be discriminated based on population responses in A1, in  
193 the task-engaged and in the passive conditions.

194

195 We first compared the mean firing-rates evoked by target and reference click trains.  
196 While some units elevated their activity for the target stimulus (Fig. 2a, left), others  
197 preferred the reference (Fig. 2a, right). Over the whole population, mean firing rates  
198 were not significantly different for target vs reference stimuli (Fig. 2b) or for low vs  
199 high rate click trains (Fig. S2a). This observation held in both passive and task-

200 engaged conditions. Discriminating between the stimuli was thus not possible on the  
 201 basis of population-averaged firing rates (see Fig. S2b).  
 202



**Fig1.** Task structure and neural encoding of click times in A1

203

**Fig 1.**

a. Structure of the click-train discrimination task and average behavior of the two animals. Each sound sequence is composed of 0.4s silence then a 1.25s long white noise burst followed by a 0.8s click train and a 0.8s silence. On each block the ferret is presented with a random number (1-7) of reference stimuli (top) preceding a target stimulus (bottom), except on catch trials with no target presentations. On blocks including a target, the animal had to refrain from licking during the final 0.4s of the trial, the no go period, to avoid a mild tail shock. (error bars are +/- sem)

b. PSTH of two example units during reference sequences in the passive and engaged state. Note that in the task-engaged state, the units show enhanced firing during the initial silent period of spontaneous activity and reduced phase locking to the stimulus.

c. Modulation index of each unit for spontaneous firing rate, spontaneous-corrected click-evoked firing rate and vector strength showing higher spontaneous firing rates and lower vector strength in the task-engaged state. The vector strength was only calculated for units firing above 1 Hz and values for both reference and target are shown. SEM error bars are not shown because not visible at this scale: 0.017, 0.037 and 0.013 respectively. (one-sample two-sided Wilcoxon signed rank test with mean 0,  $n=370, 574, 370$ ,  $zval=-8.99$ ,  $p=2.57e-19$ ;  $zval=-0.07$ ,  $p=0.94$ ;  $zval=-8.82$ ,  $p=1.16e-18$ ; \*\*\*:  $p<0.001$ ).

d. Schematic of stimulus reconstruction algorithm. Using PSTHs from half of the trials, a time-lagged filter is fitted to allow optimal reconstruction of the stimulus for each individual unit. Individual reconstructions are summed to obtain a population reconstruction (far right).

e. Stimulus reconstruction from an example session showing degraded reconstruction in the task-engaged state.

f. Mean click reconstruction in passive and engaged states.

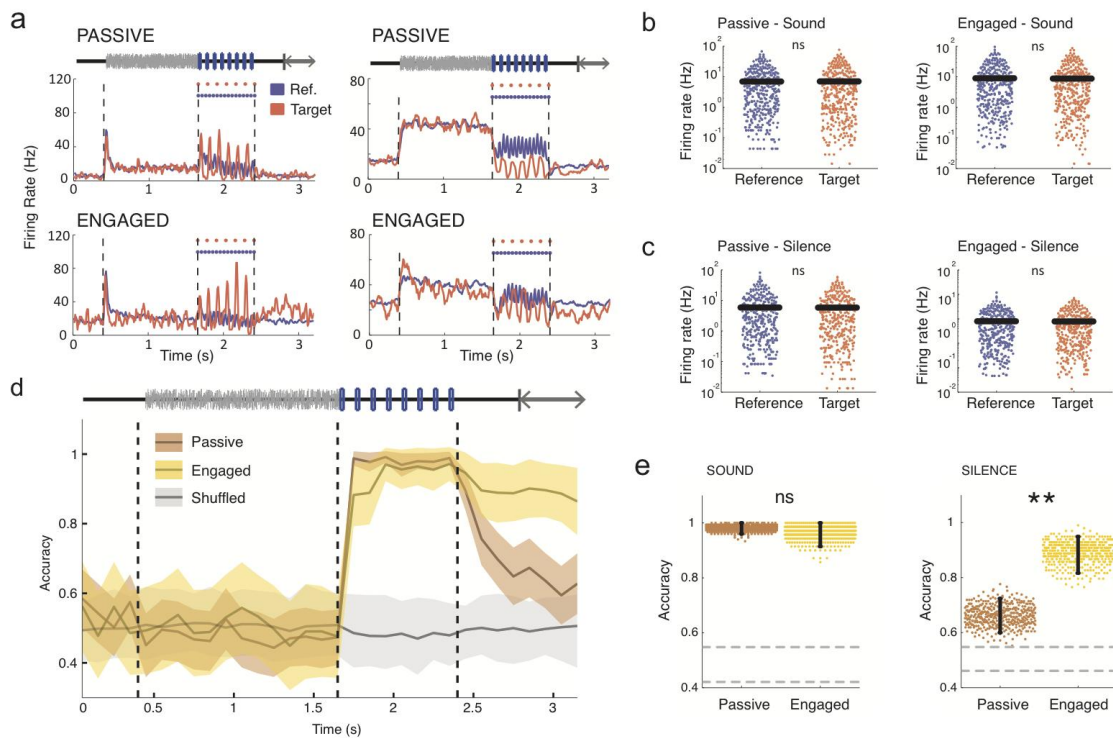
g. Modulation index of each session for stimulus reconstruction error. SEM error bar is not shown because not visible at this scale: 0.0014. (one-sample two-sided Wilcoxon signed rank test with mean 0,  $n=36$ ;  $zval=-3.4092$ ,  $p=6.51e-4$ ; \*\*\*:  $p<0.001$ ).

204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248

To take into account the heterogeneity of neural responses and quantify the ability of the whole population to discriminate between target and reference stimuli on an individual trial basis, we adopted a population-decoding approach. We used a simple, binary linear classifier that mimics a downstream readout neuron. The classifier takes as inputs the spike-counts of all the units in the recorded population, multiplies each input by a weight, and compares the sum to a threshold to determine whether a trial was a reference or a target. The weight of each unit was set based on the difference between the average spike-counts evoked by the two stimuli (Fig. S3 and Methods). This weight was therefore positive or negative depending on whether it preferred the target or reference stimulus. Different decoder weights were determined at every time-bin in the trial. The width of the time-bins (100ms) was larger than the inter-click intervals (Methods). Shorter time-bins increase the amount of noise but do not affect our main findings (Fig. S8A). Training and testing the classifier on separate trials allowed us to determine the cross-validated performance of the classifier, and therefore the ability to discriminate between the two stimulus classes based on single-trial activity in A1.

During stimulus presentation, the linear readout could discriminate target and reference stimuli with high accuracy in both passive and task-engaged conditions (Fig. 2d,e). Because the classifier performed at saturation during the sound epoch, it could be that differences between passive and active classifiers were masked by the substantial number of neurons provided to the classifiers. Decoders performing with lower numbers of neurons did not reveal any difference between the two behavioral states (Fig. S4a). Moreover this discrimination capability did not appear to be layer-dependent (Fig. S4b,c). The primary auditory cortex therefore appeared to robustly represent information about the stimulus class, independently of the decrease in the encoding of precise stimulus properties that occurs during task-engagement.

We next examined the discrimination performance during the silence immediately after stimulus offset. This silent period consisted of a 400ms interval followed by a response window, during which the animal learned to stop licking if the preceding stimulus was a target. As during the sound period, mean firing rates were not significantly different for the two types of stimuli during post-stimulus silence (Fig. 2c). Nevertheless, we found that discrimination performance between target and reference trials remained remarkably high throughout the post-stimulus silence in the task-engaged condition. In the passive condition, the decoding performance decayed during post-stimulus silence, but remained above chance level (Fig. 2d,e and Fig. S5b). The information about the stimulus class was thus maintained during the silent period in the neural activity in A1, but more strongly when the animal was actively engaged in the task. Moreover, a comparison between the decoders determined during the sound and after stimulus presentation showed that the encoding of information changed strongly between the two epochs of the trial (Fig. S6 and supplementary text).



**Fig2.** Discrimination of target and reference stimuli based on A1 activity

249  
250

**Fig 2**

a. PSTHs of two example units during reference (blue) and target (red) trials in the passive (top) and task-engaged (bottom) state. The unit on the left is target-preferring and the unit on the right is reference-preferring.

b-c. Comparison of average firing rates on a log scale in passive (left) and engaged (right) between target and reference stimuli during the sound (b) and during the post-stimulus silence (c) periods. SEM error bars are not shown because not visible at this scale. (two-sided Wilcoxon signed rank,  $n=370$ ;  $zval=0.34$ ,  $p=0.73$ ;  $zval=0.35$ ,  $p=0.79$ ;  $zval=-0.47$ ,  $p=0.64$ ;  $zval=-0.35$ ,  $p=0.73$ )

d. Accuracy of stimulus classification in passive and engaged states. In grey, chance level performance evaluated on label-shuffled trials. Error bars represent 1 std calculated over 400 cross-validations.

e. Mean classifier accuracy during the sound (left) and silence period (right) in both conditions. Gray dotted lines give 95% confidence interval of shuffled trials. Error bars represent 95% confidence intervals. ( $n=400$  cross validations;  $p=0.29$  and  $p<0.0025$ ; \*\*:  $p<0.01$ )

251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263

**Task-engagement shifts encoding towards enhanced target-detection**

We next examined in more detail the neural activity that underlies the classification performance in the two conditions. Target and reference stimuli play highly asymmetric roles in the Go/No-Go task design studied here as their behavioral meaning is totally different. As shown in Figure 1a, animals continuously licked throughout the task and only target stimuli elicited a change from this ongoing behavioral output while reference stimuli did not. We therefore sought to determine whether target- and reference-induced neural responses play similar or different roles in the discrimination between target and reference stimuli.

264 We first used dimensionality-reduction techniques to visualize the trajectories of the  
265 population activity in three dimensions (Fig. 3a, see Methods for details). The three  
266 principal dimensions were determined jointly for the passive and active data. This  
267 allowed us to visually inspect the difference in population dynamics and decoding  
268 axes between the two behavioral conditions. The average neural trajectories on  
269 reference and target trials strongly differ in the two behavioral conditions. In the  
270 passive condition, reference and target stimuli led to approximately symmetric  
271 trajectories around baseline spontaneous activity, suggesting that reference and  
272 target stimuli played essentially equivalent roles during the sound (Fig. 3a,c,d). In  
273 contrast, in the task-engaged condition, the activity evoked by reference and target  
274 stimuli became strongly asymmetric with respect to the decoding axes and the  
275 spontaneous activity (Fig. 3b,e,f).

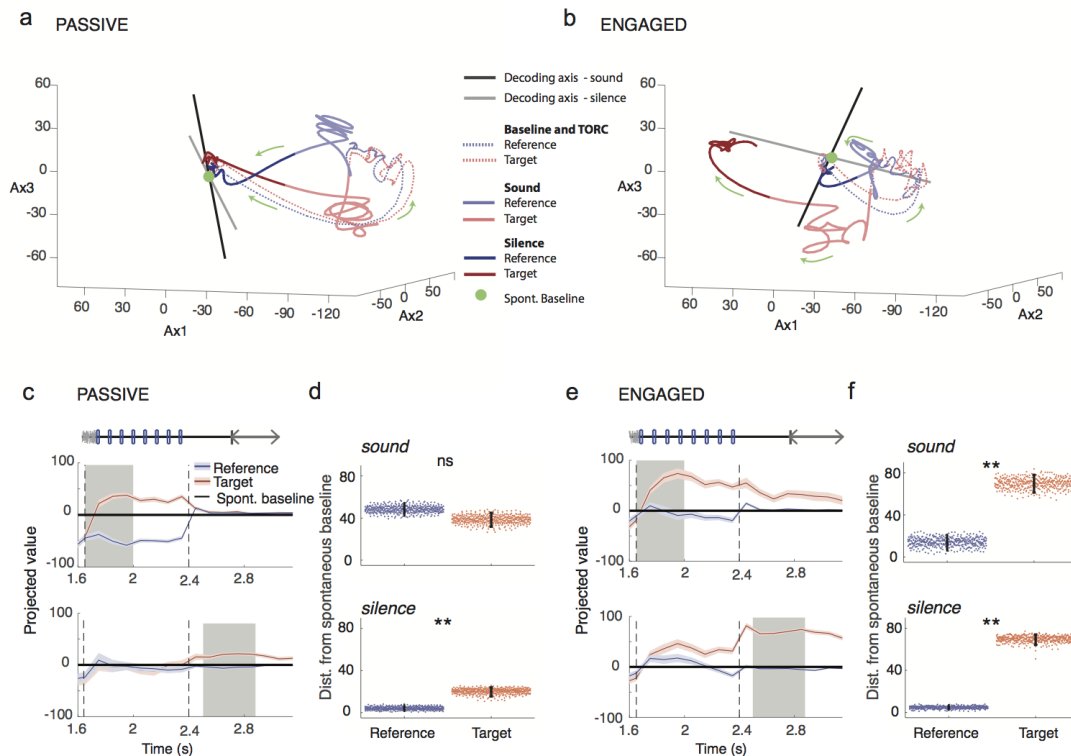
276  
277 To further characterize the change in information representation between the two  
278 conditions, we examined the average inputs from target and reference stimuli to a  
279 hypothetical readout neuron corresponding to a previously determined linear  
280 classifier. This is equivalent to projecting the trial-averaged population activity onto  
281 the axis determined by the linear classifier, trained at a given time point in the trial.  
282 This procedure sums the neuronal responses after applying an optimal set of  
283 weights. It effectively reduces the population dynamics from  $N=370$  dimensions  
284 (where each dimension represents the activity of an individual neuron) to a single,  
285 information-bearing dimension. The discrimination performance of the classifier is  
286 directly related to the distance between reference and target activity after projection,  
287 so that the projection allows us to visualize how the classifier extracts the stimulus  
288 category from the neuronal responses to the two respective stimuli. Projecting the  
289 spontaneous activity along the same axis provides moreover a baseline for  
290 comparing the changes in activity induced by the target and reference stimuli along  
291 the discrimination axis. As the encoding changes strongly between stimulus  
292 presentation and the subsequent silence (Fig. S6 and supplementary text), we  
293 examined two projections corresponding to the decoders determined during stimulus  
294 and during silence.

295  
296 As suggested by the three-dimensional visualization, the projections on the decoding  
297 axes demonstrated a clear change in the nature of the encoding between the two  
298 behavioral conditions. In the passive condition, reference and target stimuli led to  
299 approximately symmetric changes around baseline spontaneous activity (Fig. 3c,d).  
300 In contrast, in the task-engaged condition, the activity evoked by reference and target  
301 stimuli became strongly asymmetric (Fig. 3e,f). In particular, the projection of  
302 reference-evoked activity remained remarkably close to spontaneous activity  
303 throughout the stimulus presentation and the subsequent silence in the task-engaged  
304 condition. The strong asymmetry in the engaged condition, and the alignment of  
305 reference-evoked activity were found irrespective of whether the projection was  
306 performed on decoders determined during stimulus (Fig. 3e,f, top) or during silence  
307 (Fig. 3e,f, bottom). The time-courses of the two projections were however different,  
308 with target-evoked responses rising very rapidly (Fig. 3e,f top) when projected along  
309 the first axis, but much more gradually when projected along the second axis (Fig.  
310 3e,f, bottom). In both cases, however, our analysis showed that in the active  
311 condition the discrimination performance relies on an enhanced detection of the  
312 target.

313

314 The strong similarity between the projection of reference-evoked activity and the  
 315 baseline formed by the projection of spontaneous activity is not due to the lack of  
 316 responses to reference stimuli in the engaged condition. Reference stimuli do evoke  
 317 strong responses above spontaneous activity in both passive and task-engaged  
 318 conditions. However, in the task-engaged, but not in the passive condition, the  
 319 population response pattern of the reference stimuli appears to become orthogonal to  
 320 the axis of the readout unit during behavior. The strong asymmetry between  
 321 reference- and target-evoked responses is therefore seen only along the decoding  
 322 axis, but not if the responses are simply averaged over the population, or averaged  
 323 after sign correction for the preference between target and reference (Fig. S7).  
 324

325 We verified that these results are robust across a range of time bins (10ms-200ms),  
 326 allowing us to cover timescales both on the order of the click rate and much longer.  
 327 Both the increase in post-sound decoding accuracy in the engaged state and the  
 328 increased asymmetry of target/reference representation were observed at all time  
 329 scales (Fig. S8a,b).  
 330  
 331



**Fig3.** Task engagement induces shift from symmetric to asymmetric representation of target and reference stimuli

332  
 333

**Fig 3.**

*a. Population response during target and reference stimuli in the passive state along the first three components identified using GPFA (see methods) on single trial data. The session begins at the baseline (green dot), followed by the TORC presentation, (dotted line) then the click presentation of either the target and the reference sound (light red and blue respectively) and finally to the post-sound silence period (dark red and blue). Note in particular that in the passive state, the reference and target activities move away symmetrically from the baseline point given by projection of spontaneous activity.*

b. As in a, for the task-engaged state. Note that in this state, target activity makes a much larger excursion from the baseline than reference activity. The axes are the same as in panel a, as the GPFA analysis was performed jointly on passive and engaged data.

c. Projection onto the decoding axis of trial-averaged reference- and target-evoked responses for the whole neural population. A baseline value computed from pre-stimulus spontaneous activity was subtracted for each unit, so that the origin corresponds to the projection of spontaneous activity (shown by black line). Decoding axes determined during sound presentation and post-stimulus silence are respectively used for projections in the top and bottom rows. The periods used to construct the decoding axis are shaded in gray. Error bars represent 1 std calculated using decoding vectors from cross-validation. This procedure allows visualization of the distance between reference and target evoked projections (that corresponds to decoding strength) and the distance of the stimuli-evoked responses from the baseline of spontaneous activity can be interpreted as the contribution of each stimulus to decoding accuracy.

d. Distance of reference and target projections from baseline in each condition during the sound and silence period. Error bars represent 95% confidence intervals ( $n=400$  cross validations;  $p=0.15$  &  $p<0.0025$ ; \*\*:  $p<0.01$ ).

e. As in c for the engaged state.

f. As in d for the engaged state. ( $n=400$  cross validations;  $p<0.0025$  &  $p<0.0025$ ; \*\*:  $p<0.01$ ).

334

335

336

### **Encoding of stimulus behavioral meaning in A1 is independent of motor activity and reflects behavioral outcomes**

337

338

339

340

341

342

343

344

345

346

347

One simple explanation of the asymmetry between target- and reference-evoked responses could potentially be the motor-evoked neuronal discharge. Indeed, during task-engagement, the animals' motor activity was different following target and reference stimuli as the animals refrained from licking before the No-Go window following the target stimulus but not the reference stimulus (Fig. 1a). As neural activity in A1 can be strongly modulated by motor activity<sup>17</sup>, such effects could potentially account for the observed differences between target- and reference-evoked population activity.

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

To assess the role played by motor activity in our findings, we first identified units with lick-related activity. To this end, we used decoding techniques to reconstruct lick timings from the population activity, and determined the units that significantly contributed to this reconstruction by progressively removing units until licking events could not anymore be detected from the population activity. We excluded a sufficient number of neurons (10%) such that a binary classifier using the remaining units could no longer classify lick and no-lick time points as compared with random data ( $p>0.4$ ; Fig. 4a,b, see Methods). We then repeated the previous analyses after removing all of these units. The discrimination performance between target and reference trials remained high and significantly different between the passive and the task-engaged conditions during the post-stimulus silence (Fig. 4c,d), while projection of target- and reference-elicited activity on the updated decoders still showed a strong asymmetry in favor of the target (Fig. 4e,f). This indicated that the information about the behavioral meaning of stimuli was represented independently of any overt motor-related activity. In all subsequent analyses we excluded all lick-responsive neurons.

364

365

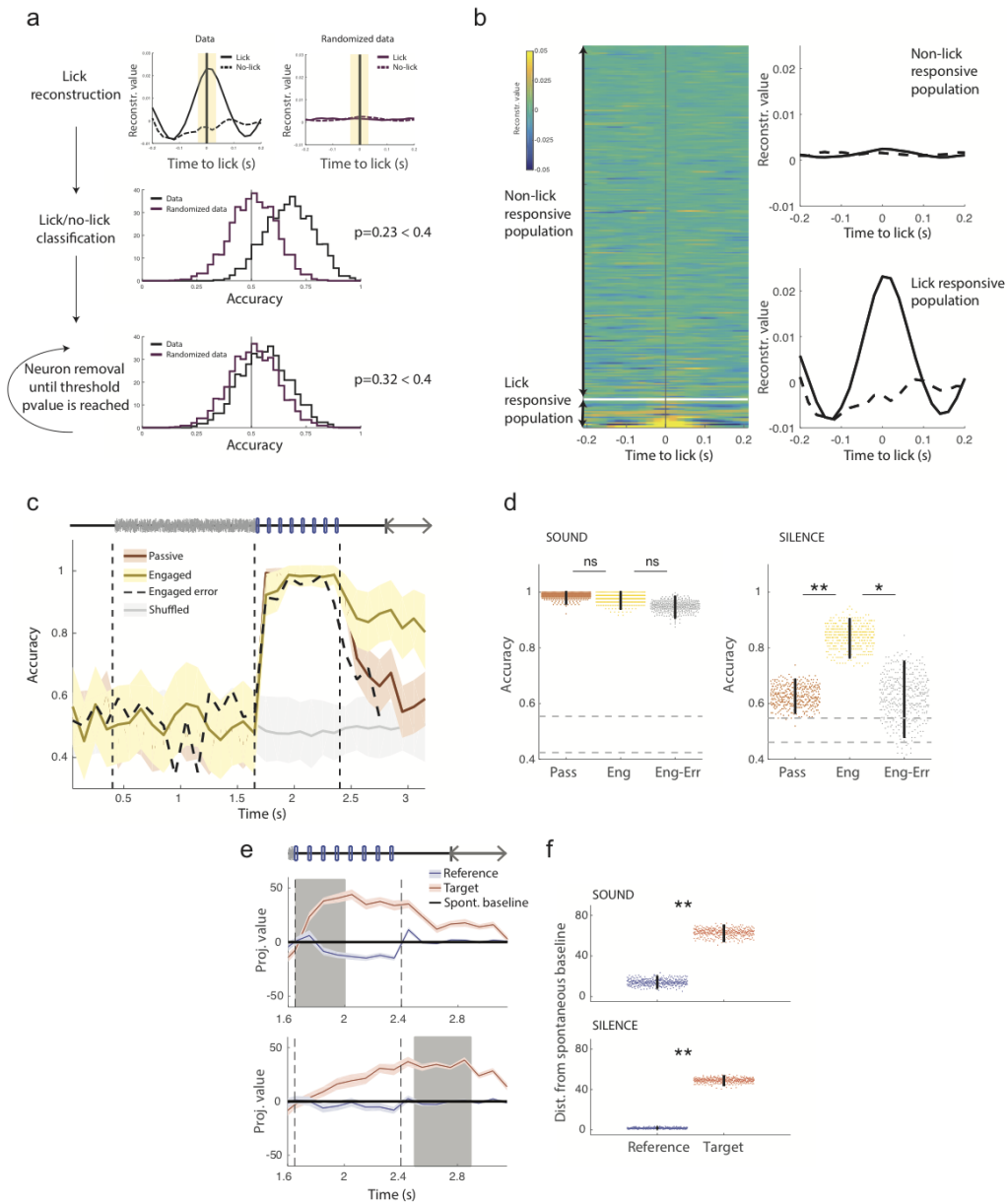
366

367

368

Although the information present in A1 during the post-stimulus silent period could not be explained by motor activity, it appeared to be directly related to the behavioral performance of the animal. To show this, we classified population activity on error trials, in which the animal incorrectly licked on target stimuli, using classifiers trained on correct trials. Error trials showed only a slight impairment of accuracy during the

369 sound presentation, but strikingly, the discrimination accuracy of the classifier during  
 370 the post-stimulus silence on these trials dropped down to the performance level  
 371 measured during passive sessions (Fig. 4c,e). This analysis therefore demonstrated  
 372 a clear correlation between the behavioral performance and the information on  
 373 stimulus category present during the silent period in A1.  
 374  
 375



376

**Fig4.** Relation between A1, motor activity and behavioural outcome

**Fig 4.**

a. Schematic of the approach used to identify lick responsive units to eliminate from population analysis. First, we reconstructed licks using optimal filters as with click reconstruction (Fig 1). To test whether this reconstruction allows to detect lick events, the filter is applied during licks and also during randomly selected time points with no licks (top left) to all units. Each event (lick or no-lick) can therefore be represented by a population vector constituted of the peak reconstruction values for all neurons. We evaluated the accuracy of classifying lick and no-lick time events using a linear decoder applied to this population vector (black distribution, middle panel). The same procedure was applied to randomized data (top right and purple distribution, middle panel) to test the significance of decoding and calculate a p-value (percentage of random data cross validations larger than real data cross validations). We then iteratively removed the best classification units (bottom plot) until the p-value was greater than 0.4 and the two distributions were indistinguishable. (see Methods for details)

b. Results of reconstruction of lick events and removal of lick units. Left shows a heatmap of average lick reconstruction for all neurons ordered by their classification weight. Right shows the average reconstruction of lick and no-lick events using units retained for population analysis (non-lick responsive) and units excluded from the population analysis (lick-responsive).

c. Accuracy of stimulus classification in passive and engaged states using only non lick-responsive units. For the engaged state both correct and incorrect trials are shown. Note that after removal of lick-responsive units, the discrimination during post-stimulus silence is still enhanced in the task-engaged state on correct trials but is low during error trials. Error bars represent 1 std calculated over 400 cross-validations.

d. Comparison of mean accuracy on passive, task-engaged correct and task-engaged error trials, during the sound (left) and post-stimulus silence periods (right). Error bars represent 95% confidence intervals. (n=400 cross validations ; sound : pass/eng p=0.22, eng/err: p=0.87; silence : pass/eng p<0.0025, eng/err: p=0.012; \*: p<0.05, \*\*: p<0.01)

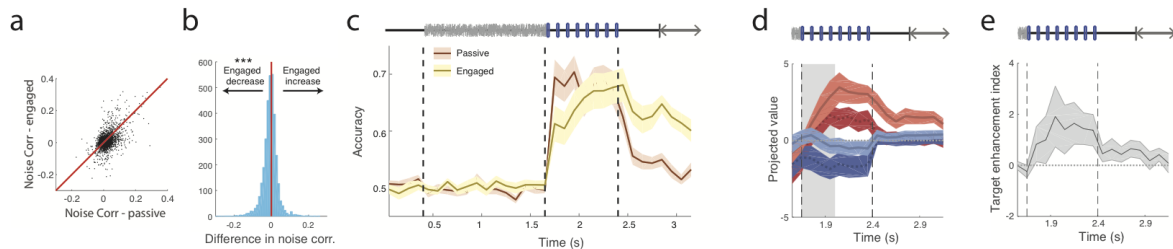
e. Projection onto the decoding axis of baseline-subtracted population vectors during the engaged condition constructed using activity of non-lick responsive units only for the reference and target stimuli. Projections are shown onto the decoding axes obtained on early sound (top) and silence periods (bottom). The periods used to construct the decoding axis are shaded in gray. A baseline value computed from pre-stimulus spontaneous activity was subtracted for each unit, so that the origin corresponds to the projection of spontaneous activity (shown by black line). Error bars represent 1 std calculated using decoding vectors from cross-validation.

f. Distance of reference and target projections from baseline in the engaged condition during the sound and silence periods. Error bars represent 95% confidence intervals (n=400 cross validations; p<0.0025 & p<0.0025; \*\*: p<0.01).

377  
378

379 Another aspect of neural activity that can be expected to change with task  
380 engagement is correlations between pairs of neurons. Our analysis so far has  
381 focused on the structure of population responses to external stimuli (signal  
382 correlations) but pairs of neurons display trial-to-trial fluctuations in activity (noise  
383 correlations) that can affect the population ability to encode information<sup>32,33</sup>. We  
384 found that task engagement decreased noise correlations on average (Fig. 5a,b; Fig.  
385 S9a), compatible with previous observations that attention reduces noise correlations  
386<sup>34</sup>. Across the population, the range of changes was however very broad. To  
387 determine the influence of noise correlations on the population level, we repeated our  
388 analysis on simultaneously recorded data, using a modified linear decoder that takes  
389 noise correlations into account (the Fisher discriminant, see Methods). Our main  
390 findings appeared not to be sensitive to noise correlations. We were able to decode  
391 with high accuracy stimulus identity in passive and engaged states and observed an  
392 increase of stimulus memory in the engaged state as before (Fig. 5c). Projection onto  
393 this adjusted decoding axis showed a similar enhanced target representation in the  
394 engaged state, with the reference response lying along the projected baseline activity  
395 (Fig. 5d,e). Projection of responses using the linear classifier with and without taking  
396 noise correlations into account are strikingly similar across a range of timebins (Fig.

397 S9b,c). Finally, a finer examination of the change between passive and engaged  
398 conditions showed that, contrary to previous observations<sup>35</sup>, noise correlations were  
399 most strongly reduced for pairs of neurons with opposite stimulus preference in our  
400 data set (Fig. S10b,c), which is expected to impair decoding of information (Fig.  
401 S10a).  
402



403 **Fig 5.** Task-induced changes in stimulus representation are independent of changes in noise correlations

**Fig 5.**

a. Comparison of noise correlations between pairs of neurons in the passive and engaged state. Red line indicates identity line.

b. Histogram of correlation changes between the engaged and passive states showing a shift to lower values in the engaged state despite highly heterogeneous behavior across the population. (two-sided Wilcoxon signed rank,  $n=3361$  pairs;  $zval=10.33$ ,  $p=4.9E-25$ ,  $***:p<0.001$ )

c. Accuracy of stimulus classification in passive and engaged states using simultaneously recorded, non lick-responsive units and applying a decoding vector corrected for noise correlations. Note that the increase in decoding accuracy during the silent period in the engaged state is still clearly visible. Error bars represent s.e.m over  $n=15$  sessions.

d. Projection onto the decoding axis determined during the sound period of trial-averaged reference (blue) and target (red) activity during the passive (dark colors) and the active (light colors) sessions. A baseline value computed from pre-stimulus spontaneous activity was subtracted for each neuron, so that the origin corresponds to the projection of spontaneous activity (shown by black line). Note that the target-driven activity lies further from the baseline in the active state and the reference-driven activity lies closer to baseline. The period used to construct the decoding axis is shaded in gray. Error bars represent s.e.m over  $n=15$  sessions

e. Index of target enhancement induced by task engagement based on projections using the decoding axis determined during the sound. This value is positive if projected target activity is enhanced in the active state and projected reference activity is reduced. Error bars represent s.e.m over  $n=15$  sessions.

404

405

406 **Mechanisms underlying the asymmetric, target-driven encoding during task-**  
407 **engagement**

408

409 The previous analyses of population activity have shown that task engagement  
410 induces an asymmetric encoding, in which the activity elicited by reference stimuli  
411 becomes similar to spontaneous background activity when seen through the  
412 decoder. Two different mechanisms can potentially contribute to this shift between  
413 passive and engaged conditions: (i) the spontaneous activity changes between the

414 two behavioral states such that its projection on the decoding axis becomes more  
415 similar to reference-evoked activity; (ii) stimulus-evoked activity changes between the  
416 states, inducing a change in the decoding axis and in the projections. In general, both  
417 mechanisms can be expected to contribute and their effects can be separated during  
418 different epochs of the trial.

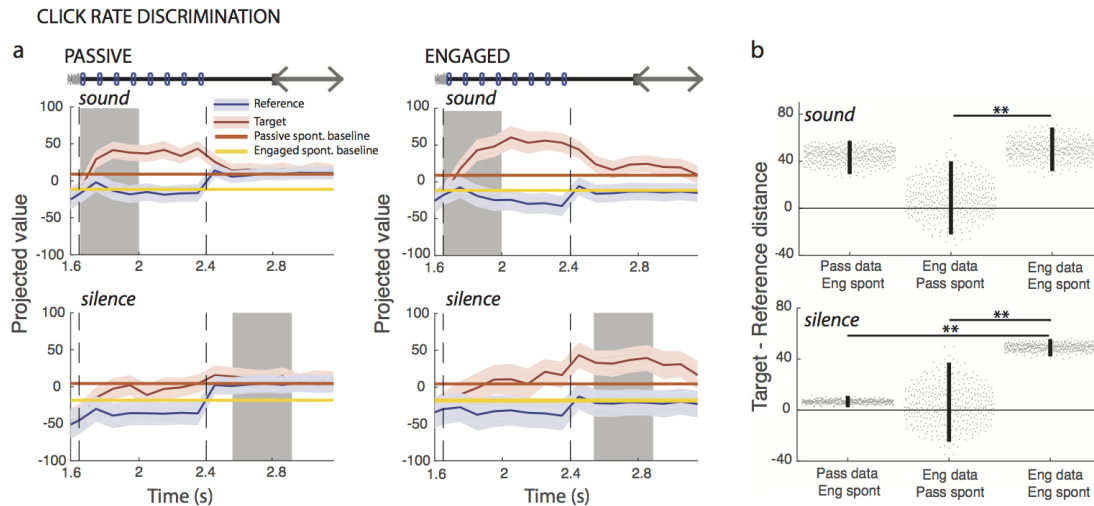
419  
420 To disentangle the effects of the two mechanisms, we chose a fixed decoding axis,  
421 and projected on the same axis the stimulus-evoked activity from both passive and  
422 engaged conditions. We then compared the resulting projections with projections of  
423 both passive and engaged spontaneous activity. We performed this procedure  
424 separately for decoding axes determined during sound and silence epochs.

425  
426 Figure 6a (top) illustrates the projections along the decoding axis determined during  
427 the sound epoch in the engaged condition. Comparing the passive responses with  
428 the passive and engaged spontaneous activity revealed that the projection of passive  
429 reference-evoked activity was aligned during sound presentation with the projection  
430 of engaged, but not passive spontaneous activity (Fig. 6a top left). A similar  
431 observation held for the engaged responses throughout the sound presentation  
432 epoch (Fig. 6a top right). These projections remained similar regardless of whether  
433 the decoding axes were determined during the passive or the engaged conditions, as  
434 these two axes largely share the same orientation (Fig. S6e). Altogether, these  
435 results indicate that the change in spontaneous baseline activity during task  
436 engagement is sufficient to explain the strongly asymmetric, target-driven response  
437 observed early in the trial during sound presentation (Fig. 6b top).

438  
439 However, we reached a different conclusion when we examined the activity during  
440 the post-stimulus silence (Fig. 6a bottom). Repeating the same procedure as above,  
441 but projecting on the decoding axis determined during the post-stimulus silence  
442 revealed that the shift in spontaneous activity alone was not able to account for the  
443 asymmetry of the projected responses during the post-stimulus silence (Fig. 6b  
444 bottom). The target-driven, asymmetrical projections observed during this trial epoch  
445 therefore relied in part on a change in stimulus-evoked responses.

446  
447 All together, we found that the changes in baseline spontaneous activity induced by  
448 the task engagement are key in explaining the enhancement of the target-driven,  
449 asymmetric encoding during sound presentation. As described in the above, the  
450 encoding axis during sound presentation is not drastically affected by task  
451 engagement. Instead, it is the population spontaneous activity that aligns with the  
452 reference-elicited activity with respect to the decoding axis. This observation in  
453 particular provides an additional argument against the possibility that the appearance  
454 of an asymmetrical representation is due to the asymmetrical motor responses to the  
455 two stimuli. Rather, the asymmetry is geometrically explained by baseline changes  
456 that precede stimulus presentation, and reflects the behavioral state of the animal.

457  
458  
459  
460  
461



**Fig 6.** Shift in spontaneous activity contributes to change in asymmetry

462

**Fig 6.**

Note that all analysis in this figure is done after excluding lick-responsive units in A1 as described in Fig 4.

a. Projection onto the engaged decoding axis of reference- and target-evoked activity in the passive (left column) and engaged state (right column). Decoding axes determined during sound presentation and post-stimulus silence are respectively used for projections in the top and bottom rows. This figure differs from Fig 3c in which the spontaneous activity is subtracted before projection. Passive and engaged spontaneous activities after projection are shown by continuous lines. Error bars represent 1 std calculated using decoding vectors from cross-validation (n=400).

b. Comparison of reference/target asymmetry for evoked responses in different states compared to different baselines given by passive or engaged spontaneous activity. Reference/target asymmetry is the difference of the distance of reference and target projected data to a given baseline. We examine three cases: (i) passive evoked responses, distances calculated relative to engaged spontaneous activity; (ii) engaged evoked responses, distances calculated relative to passive spontaneous activity; (iii) engaged evoked responses, distances calculated relative to engaged spontaneous activity. These values are shown during the sound (top) and the silence (bottom). In all three cases, the engaged decoding axis was used for projections. Decoding axes determined during sound presentation and post-stimulus silence are respectively used for projections in the top and bottom rows.

Error bars represent 95% confidence intervals (n=400 cross validations; sound:  $p(\text{col1}, \text{col3})=0.29$  &  $p(\text{col2}, \text{col3}) < 0.0025$ ; silence:  $p(\text{col1}, \text{col3}) < 0.0025$  &  $p(\text{col2}, \text{col3}) < 0.0025$ ; \*\*:  $p < 0.01$ ).

463

464

465

466

**Sustained, target-driven, and behaviorally-gated responses of single cells in frontal cortex parallel population encoding in A1**

467

468

469

470

471

472

473

474

475

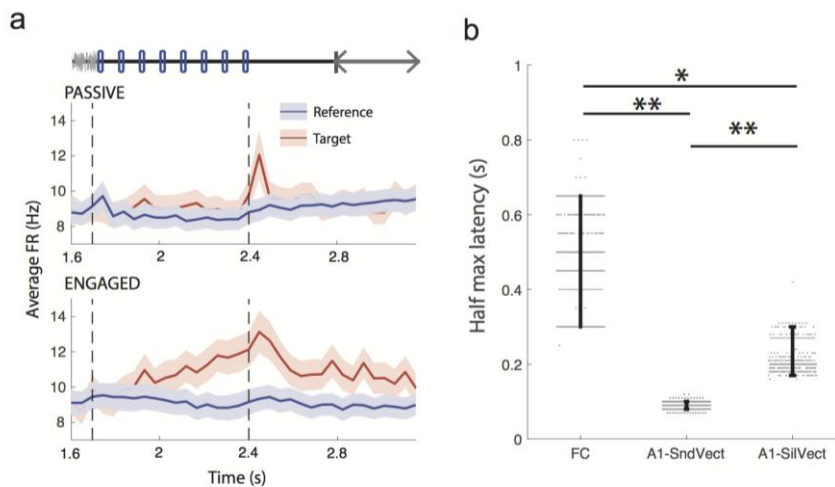
476

477

The pattern of activity resulting from projecting reference- and target-elicited A1 activity on the linear readout is strikingly similar to previously published activity recorded in the dorsolateral frontal cortex (dlFC) of behaving ferrets performing similar Go/No-Go tasks (tone detect and two-tone discrimination in <sup>36</sup>). We therefore compared in more detail A1 activity with activity recorded in dlFC during the same click-rate discrimination task. When the animal was engaged in the task, single units in dlFC encoded the behavioral meaning of the stimuli by responding only to target stimuli, but remaining silent for reference stimuli (Fig. 6a bottom panel). Target-induced responses were moreover observed well after the end of the stimulus presentation, allowing for a maintained representation of stimulus category. The strong asymmetry of single-unit responses in dlFC clearly resembles the activity

478 extracted from the A1 population by the linear decoder (Fig. 3 and 4). This suggests  
479 that the target-selective responses in the dIFC that reflect the cognitive decision  
480 process could in part be thought of as a simple readout of information already  
481 present in the population code of A1.

482  
483 To further examine the relationship between dIFC single-unit responses and  
484 population activity in A1, we next compared the time course of the projected target-  
485 elicited data in A1 (Fig. 3e) and the population-averaged target-elicited neuronal  
486 activity in dIFC (Fig. 7a bottom panel) during active sessions. As mentioned above,  
487 the optimal decoding axes for A1 activity changes between the stimulus presentation  
488 epoch and the silence that follows (Fig. S6). The time-course of the projected A1  
489 activity depends strongly on the axis used for the projection. When projecting on the  
490 axis determined during stimulus presentation, the target-elicited response in A1 was  
491 extremely fast ( $0.08s \pm 0.009$  std) compared to the much longer response latency in  
492 the population-averaged response of dIFC neurons ( $0.48s \pm 0.12$  std) (Fig. 7b). In  
493 contrast, when projecting on the axis determined during post-stimulus silence, the  
494 target-elicited response in A1 was slower ( $0.21s \pm 0.03$  std) and closer to the  
495 population-averaged response in the dIFC (note that a fraction of individual units in  
496 dIFC display a very fast responses not reflected in the population average, see Fritz  
497 et al. 2010). Our analyses therefore identified two contributions to target-driven  
498 population dynamics in A1, a fast component absent in population-averaged dIFC  
499 activity and a slower component similar to population-averaged activity in dIFC, thus  
500 pointing to a possible contribution of an A1-FC loop that could be engaged during  
501 auditory behavior.



**Fig7.** Persistent, asymmetric response to target and reference stimuli in frontal cortex

502

**Fig 7.**

Note that all analysis in this figure is done after excluding lick-responsive units in A1 as described in Fig 4.  
a. Average PSTHs of all frontal cortex units in response to target and reference stimuli in both passive and engaged conditions. Note that the response to the target in the task-engaged state is very clear and appears late during the sound. Error bars: s.e.m over all units (n=102)  
b. Latency to half-maximum response for frontal cortex (for average PSTHs) and primary auditory cortex (for projected target-elicited data) in the task-engaged state. For the auditory cortex, data is projected either on the sound decoding vector or the silence decoding vector. Error bars represent 95% confidence intervals. (400 cross-validations.  $p < 0.0025$ ,  $p < 0.0025$  &  $p = 0.011$ ; \*\*:  $p < 0.01$ , \*:  $p < 0.05$ ).

503

504

505 **Enhanced representation of target stimuli in A1 is a general feature of auditory**  
506 **Go/No-Go tasks**

507  
508 To determine whether the task-related increase in asymmetry between target and  
509 reference was a more general feature of primary auditory cortex responses during  
510 auditory discrimination, we applied our population analysis to other datasets collected  
511 during different tasks. All of these tasks used Go/No-Go paradigms (see Fig. S11a,e,i  
512 and Methods), in which the animals were presented with a random number of  
513 references followed by a target stimulus. In these different datasets, animals were  
514 required to discriminate noise bursts vs. pure tones (tone detect tasks), or categorize  
515 pure tones drawn from low, medium or high-frequency ranges (frequency range  
516 discrimination task). Contrasting datasets were obtained from two groups of ferrets  
517 that were separately trained on approach and avoidance versions of the same tone  
518 detect task. These two behavioral paradigms used exactly the same stimuli under  
519 two opposite reinforcement conditions<sup>30</sup>, requiring nearly opposite motor responses  
520 (Fig. S11a,e). A crucial feature shared by all these tasks lies in the fact that the  
521 behavioral response to the target stimulus always required a behavioral change  
522 relative to sustained baseline activity. More specifically the target was the No-Go  
523 stimulus in negative reinforcement tasks and required animals to *cease* ongoing  
524 licking, whereas the target was the Go stimulus in the positive reinforcement task and  
525 required animals to *begin* licking in a non-lick context. In all of the analyses, lick-  
526 related neurons were removed using the approach outlined earlier.

527  
528 Performing the same analyses on all tasks showed that projections of target- and  
529 reference-evoked activities in passive conditions contained a variable degree of  
530 asymmetry in the sound and silence epochs. However, in all tasks we found that  
531 task-engagement leads an enhancement of target-driven encoding during sound  
532 (Fig. 8a,b;e,f; i,j; m,n). As previously described for the rate discrimination task (Fig. 3  
533 and 4e), target projections more strongly deviated from baseline than projections of  
534 reference stimuli in the engaged condition. Moreover, for three of the four tasks we  
535 examined, enhancement of target representations was not observed at the level of  
536 population-averaged responses, but only in the direction determined by the decoder  
537 (Fig. 8b,f,j,n). During the post-sound silence, decoding accuracy quickly decayed in  
538 both passive and engaged states, but remained above chance (Fig. S11c,g,k). As in  
539 the click-train detection task, decoding accuracy relied on a different encoding  
540 strategy than the sound period (Fig. S11d,h,l), and the asymmetry during the post-  
541 sound silence was high both in passive and engaged conditions (Fig. S12).

542  
543 Comparison of appetitive and aversive versions of the same task is particularly  
544 revealing as to which type of stimulus was associated with enhanced representation  
545 in the engaged state. In the appetitive version of the tone detect task, ferrets needed  
546 to refrain from licking on the reference sounds (No-Go) and started licking the water  
547 spout shortly after the target onset (Go) (Fig. S11e), whereas in the aversive  
548 (conditioned avoidance) paradigm they had to stop licking after the target sound (No-  
549 Go) to avoid a shock (Fig.S11a). It is important to note that although the physical  
550 stimuli presented to the behaving animals were identical in both tone detect tasks,  
551 the associated motor behaviors of the animals are nearly opposite. Projection of task-  
552 engaged A1 population activity reveals a target-driven encoding (compare right  
553 panels of Fig. 8f,j with Fig. 8l,j), irrespective of whether the animal needed to refrain  
554 from or to start licking to the target stimulus. This shows that the common feature of

555 stimuli that are enhanced after projection onto the decoding axis is that they are  
556 associated with a change of ongoing baseline behavior.

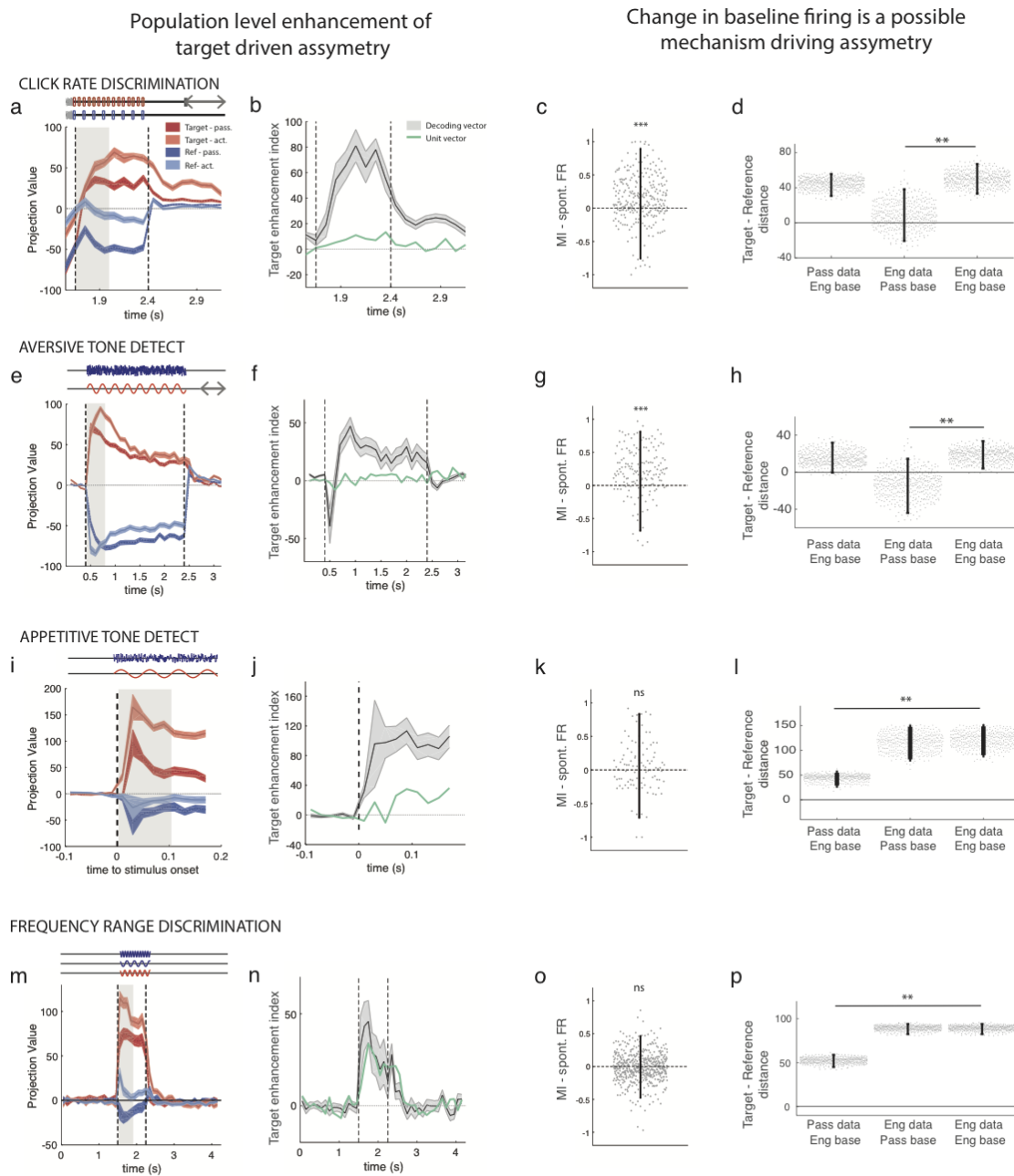
557  
558 This range of behavioral paradigms provides additional arguments against the  
559 described changes in activity being solely due to correlates of licking activity. Firstly,  
560 we observed enhanced target-driven encoding in both the appetitive and aversive  
561 tone-detect paradigms, even though the licking profiles were diametrically opposite to  
562 each other. Secondly, comparing the projections of the population activity in the  
563 approach tone detect task with the click rate discrimination task reveals a strong  
564 similarity in the temporal pattern of asymmetry observed during task engagement. In  
565 less than 100 ms, projection of target-elicited activity reached its peak in both  
566 paradigms (Fig. 8a,i), although the direction and time course of the licking responses  
567 were reversed, with a fast decline in lick frequency for the click rate discrimination  
568 task (Fig. 1a), versus a slow increase for the tone detect (Fig. S11e left panel). Last,  
569 although the results are more variable partly due to low decoding performance, we  
570 observed target-driven encoding during the post-stimulus silence in the passive state  
571 (Fig. S12) although ferrets were *not* licking during this epoch. The points listed here  
572 are again in agreement with a representation of the stimulus' behavioral  
573 consequences, independent of the animal motor response.

574  
575 As pointed out in the case of the click rate discrimination task, the enhancement of  
576 target representation in the engaged condition can rely on two different mechanisms,  
577 a shift in the spontaneous activity or a shift in stimulus-evoked activity. We therefore  
578 set out to tease apart the respective contributions of the two mechanisms in this  
579 novel set of tasks. As in Fig. 6, we compared the distance of target and reference  
580 passive and engaged projections to either engaged or passive baseline activities.  
581 Out of the three additional datasets, we observed an increase in spontaneous firing  
582 rates only in the aversive tone detect task (Fig. 8g). In this latter paradigm, task-  
583 induced modulations of spontaneous activity patterns explained the change in  
584 asymmetry during sound presentation, similar to what was observed in the click rate  
585 discrimination task (compare Fig. 8d and 8h). The other two tasks showed no global  
586 change of spontaneous firing rate (Fig. 8k,o), and consequently, during the task  
587 engagement, the enhancement of the target representation was solely due to the  
588 second mechanism, the changes in the target-evoked responses themselves  
589 (Fig.8l,p). During the silence, we observed as previously for the click-rate  
590 discrimination that the increase in asymmetry relied only on the second mechanism  
591 (Fig. S11).

592  
593 Taken all together, population analysis on four different Go/No-Go tasks revealed an  
594 increase of the encoding in favor of the target stimulus as a general consequence of  
595 task-engagement on A1 neural activity. Viewing activity changes in this light allowed  
596 us to interpret the previously observed changes in spontaneous activity as one of two  
597 possible mechanisms underlying this task-induced change of stimulus representation  
598 in A1 population activity.

599  
600  
601  
602  
603  
604

605



**Fig8.** Asymmetric encoding of target and reference stimuli in a range of auditory discrimination tasks

606

**Figure 8.**

Each line of four panels represent the same analysis for all four tasks, statistics are given in order of appearance in the figure: click rate discrimination, aversive tone detect, appetitive tone detect, frequency range discrimination.

a,e,l,m Projection onto the decoding axis determined during the sound period of trial-averaged reference (blue) and target (ref) activity during the passive (dark colors) and the active (light colors) sessions. A baseline value computed from pre-stimulus spontaneous activity was subtracted for each neuron, so that the origin corresponds to the projection of spontaneous activity (shown by black line). Note that the target-driven activity is further from the baseline in the active state and the reference-driven activity is closer. The periods used to construct the decoding axis are shaded in gray. Error bars represent 1 std calculated using decoding vectors from cross-validation (n=400).

b,f,j,n Index of target enhancement induced by task engagement based on projections using the decoding axis determined during the sound. In green same index instead giving the same weight to all units. The difference between the green and black curved indicates that the change in asymmetry induced by task engagement cannot be detected using the population averaged firing rate alone. Error bars represent 1 std calculated using decoding vectors from cross-validation (n=400).

c,g,k,o Modulation index of each unit for spontaneous firing rate after exclusion of lick-related units. Error bars are 95% C.I. (one-sample two-sided Wilcoxon signed rank test with mean 0, n=277, zval=6.35, p=2.1e-10; n=161, zval=7.22, p=5.4e-13; n=99, zval=1.01, p=0.30; n=520, zval=-0.78, p=0.47; \*\*\*: p<0.001).

d,h,l,p Comparison of reference/target asymmetry for evoked responses in different states compared to different baselines given by passive or engaged spontaneous activity. Reference/target asymmetry is the difference of the distance of target and reference projected data to a given baseline. We examine three cases: (i) passive evoked responses, distances calculated relative to engaged spontaneous activity; (ii) engaged evoked responses, distances calculated relative to passive spontaneous activity; (iii) engaged evoked responses, distances calculated relative to engaged spontaneous activity. In all three cases, the engaged decoding axis was used for projections. Error bars represent 95% confidence intervals (n=400 cross validations; p(col1,col3)=0.29 & p(col2,col3)<0.0025; p(col1,col3)=0.38 & p(col2,col3)<0.0025; p(col1,col3)<0.0025 & p(col2,col3)=0.16; p(col1,col3)<0.0025 & p(col2,col3)=0.92; \*\*: p<0.01).

607

608

## DISCUSSION

609

610 In this study, we examined population responses in the ferret primary auditory cortex  
611 during auditory Go/No-Go discrimination tasks. Comparing responses between  
612 sessions in which animals passively listened and sessions in which animals actively  
613 discriminated between stimuli, we found that task-engagement induced a shift from a  
614 sensory-driven to an asymmetric, target enhanced, representation of the stimuli,  
615 highly similar to the type of activity observed in dorsolateral frontal cortex during  
616 engagement in the same task. This enhanced representation of target stimuli was  
617 found in a variety of discrimination tasks that shared the same basic Go/No-Go  
618 structure, but used a variety of auditory stimuli and reinforcement paradigms.

619

620 In the click rate discrimination task that we analyzed first, the sustained asymmetric  
621 stimulus encoding in A1 was only observed in the engaged state (Fig. 3). One  
622 possible explanation is that this encoding scheme relied on corollary neuronal  
623 discharges related to licking activity. However there are several factors that argue  
624 against this interpretation. Firstly, we adopted a stringent criterion for the exclusion  
625 from the analysis of all units whose activity was correlated with lick events (Fig. 4).  
626 After removing lick-responsive units from the analysis the results remained  
627 unchanged, indicating the absence of a direct link between licking and the observed  
628 asymmetry in the encoding. Furthermore, the large differences in the lick profiles  
629 between the different tasks were not in line with the remarkably conserved target-  
630 driven projections of population activity across tasks and reinforcement types,  
631 supporting a non-motor nature of the stimulus encoding in A1 (Fig. 8b,f,j,n). Finally,

632 the role of baseline shifts due to the change in spontaneous activity in two more tasks  
633 further argues against a purely motor explanation of the observed asymmetry (Fig. 6  
634 and Fig. 8a) since the spontaneous activity occurs during epochs that preceded  
635 stimulus presentation and behavioral changes. Altogether, while the different lines of  
636 evidence exposed above make an interpretation in terms of motor activation unlikely,  
637 ultimately a different type of behavioral report, such as one using similar responses,  
638 would help fully rule out this possibility.

639  
640 Our analyses show that the target-driven encoding scheme during task engagement  
641 is neither purely sensory nor purely motor, but instead argue for a more abstract,  
642 cognitive representation of the stimulus behavioral meaning in A1 during task  
643 engagement. As the target stimulus was associated with an absence of licking in the  
644 tasks under aversive conditioning, one possibility could have been that the A1  
645 encoding scheme was contrasting the only stimulus associated with an absence of  
646 licking (No-Go) against all other stimuli (Go). This lick/no-lick encoding was however  
647 not consistent with the tone detect task under appetitive reinforcement, in which the  
648 target stimulus was a Go signal for the animal. We thus suggest that A1 encodes the  
649 behavioral meaning of the stimulus by emphasizing the stimulus requiring the animal  
650 to change its behavioral response, i.e. the target stimuli in the different tasks we  
651 examined. However, our data do not allow us to conclude whether this behavioral  
652 meaning corresponds to the encoding of the stimulus-action association, or the  
653 animal's decision, or the output motor command leading to a change in behavioral  
654 response and it would be interesting to perform similar analyses in tasks more  
655 specifically designed to tease apart these different possible interpretations.

656  
657

### 658 **Relation to previous studies**

659

660 A series of previous studies found that task-engagement strongly influences  
661 responses in the primary auditory cortex, in some cases sharpening stimulus  
662 representation<sup>26–28,37</sup>, in others leading to a suppression of sensory responses<sup>14</sup>, as  
663 was also observed during locomotion<sup>17,18</sup>. While some studies observed signatures  
664 of decision-related activity in A1<sup>11,38</sup>, none has hitherto reported the strong  
665 representation of behavioral meaning described here in the population code.

666

667 The majority of previous studies concentrated on single-neuron or LFP activity. In  
668 contrast, our results critically rely on population-level analyses<sup>39–42</sup>, and in particular,  
669 on linear decoding of population activity. This is a simple, biologically-plausible  
670 operation that can be easily implemented by a neuron-like readout unit that performs  
671 a weighted sum of its inputs. The summed inputs to this hypothetical read-out unit  
672 showed that Go and No-Go stimuli elicited inputs symmetrically distributed around  
673 spontaneous activity in the passive state. In contrast, in the task-engaged state, only  
674 target stimuli, which required an explicit change in ongoing behavior, led to an output  
675 different from spontaneous activity, once passed through the readout unit. This  
676 switch from a more symmetric, sensory-driven to an increasingly asymmetric, target-  
677 driven representation was not clearly apparent if single-neuron responses were  
678 simply averaged or normalized (Fig. S7, 7b,f,j,n), but instead relied on a population  
679 analysis in which different units were assigned different weights by projecting  
680 population activity on the decoding axis. Note that the weights were not optimized to  
681 maximize the asymmetry between Go and No-Go stimuli, but rather the

682 discrimination between them. The shift towards a more asymmetric representation of  
683 the behavioral meaning of stimuli is therefore an unexpected but important by-  
684 product of the analysis.

685  
686 From a population-decoding viewpoint, task-engagement induced a shift towards an  
687 enhanced representation of target stimuli class in all the tasks we considered.  
688 However, considering these same effects from a less elaborate sensory coding view,  
689 they appear to be quite varied and to depend on the details of the stimuli. Thus, in  
690 the tone-detection task, previous studies reported that task-engagement enhanced  
691 the representation of the relevant tone frequency in a negative reinforcement  
692 paradigm<sup>26–28</sup>, and caused a suppression at the tone frequency during the appetitive  
693 version of the task<sup>30</sup>. In the click-discrimination task, task-engagement led to  
694 decreased temporal fidelity in the representation of click times, the main sensory  
695 features of the stimuli (see Fig. 1 and<sup>14</sup>). These varied results, however, are unified  
696 by a shift to a representation of the behavioral meaning of stimuli. Our findings  
697 therefore provide a possible way to reconcile the diverse effects described earlier.

### 698 699 **Possible implication of an A1-FC loop during task engagement**

700  
701 Recordings performed in dorsolateral frontal cortex (dlFC) in the ferret during tone  
702 detection<sup>36</sup> showed that, when the animal is engaged in the task, dlFC single units  
703 encode the abstract behavioral meaning of the stimuli by responding only to target  
704 stimuli (that require a change in the ongoing behavioral output) but remain silent for  
705 reference stimuli. Remarkably, projections of reference- and target-elicited A1 activity  
706 on the linear readout showed the same type of target-specific patterns of activity.  
707 Several possible mechanisms could account for these similarities of representations  
708 in A1 and dlFC. Here we propose that, during task engagement, sound evoked  
709 activity in A1 triggers activity in dlFC, which then subsequently feeds back top-down  
710 inputs to A1 that may underlie the sustained activity pattern found during post-  
711 stimulus silence.

712 Very early in the trial, the asymmetric encoding is already fully present in A1 (as early  
713 as 100ms in the rate discrimination task for instance; Fig. 3e top panel). At this point  
714 in time, dlFC does show some target-selective responses that increase over time  
715 (Fig. 7a). This suggests the presence of a feed-forward mechanism early in the trial,  
716 by which A1 may be feeding higher-order auditory cortex and FC with a pattern of  
717 neuronal responses encoding the behavioral meaning of the stimulus. Our results  
718 show that this early task-induced change in the representation in A1 relies on a shift  
719 of spontaneous activity at the population level that may be due to tonic top-down or  
720 neuromodulatory inputs during task engagement<sup>43,44</sup>. The presence of a dynamic  
721 balance characterizes interactions between A1 and dlFC has been previously shown  
722 by changes in Granger causality and effective connectivity during behavioral state  
723 transitions<sup>45</sup>.

724  
725 As the trial progresses, the encoding in A1 progressively shifts (Fig. S6). Activity  
726 projected on the late decoding vector (Fig. 3e bottom panel) shows a progressive  
727 buildup similar to the activity observed in dlFC (Fig. 7). The late stimulus encoding,  
728 during the later phase of the click trains and the subsequent post-stimulus silence  
729 (Fig. 3e bottom panel) may thus be gradually engaging stronger top-down inputs from  
730 the dlFC-A1 network loop. The persistent encoding of stimuli identity could therefore

731 rely on a stimulus-specific top-down input from frontal areas. Although direct  
732 connections from dlFC to A1 have not been identified in ferrets, several recent  
733 studies have identified direct inputs from the rodent motor cortex<sup>17</sup>, the rodent  
734 orbitofrontal cortex<sup>46,47</sup> and the secondary auditory areas<sup>48</sup> (ferret posterior  
735 ectosylvian gyrus) to A1. Altogether, while the comparison of time-course of activity in  
736 A1 and dlFC suggest that the recruitment of the A1-FC loop is a plausible  
737 interpretation of our results, more direct evidence is needed to establish this  
738 mechanism.

739

## 740 **Projection to the read-out null space as a mechanism for target detection in A1**

741

742 Our analysis suggests a novel population readout mechanism for extracting  
743 behaviorally relevant information from A1 while suppressing other, irrelevant sensory  
744 information: in the task-engaged state, irrelevant sensory inputs (reference stimuli)  
745 elicit changes of activity that are orthogonal to the read-out axis and therefore cannot  
746 be distinguished from spontaneous activity. This mechanism is reminiscent of the  
747 mechanism proposed for movement preparation in motor cortex<sup>49</sup>, where  
748 preparatory neural activity lies in the null space of the motor readout, i.e. the space  
749 orthogonal to the read-out of the motor command, and therefore does not generate  
750 movements. In our case, the readout is task-dependent, as it presumably depends on  
751 the performed discrimination task. We showed that the A1 activity in the engaged  
752 condition rearranges so that the difference between spontaneous activity and  
753 reference-elicited activity lies in the null space of the readout, which is therefore only  
754 activated by target stimuli. This rearrangement can be implemented either by a  
755 change of reference-elicited activity or by a change of spontaneous activity. In two of  
756 the examined tasks, click-discrimination and aversive tone detection, we found that  
757 the rearrangement of population activity relied mostly on the change in population  
758 spontaneous activity in the engaged condition. Strikingly, these two tasks were  
759 performed by the same ferrets, which were trained to switch between the two tasks in  
760 the same session. In the two other tasks, reference-elicited activity in the passive  
761 condition were already aligned with the passive spontaneous activity when projected  
762 on the active decoder, suggesting that learning these behavioral tasks may have  
763 profoundly reshaped stimulus-evoked activity. Our results therefore suggest that  
764 task-dependent shaping of spontaneous activity can allow the primary auditory cortex  
765 to encode the behavioral meaning of stimuli in a task-relevant, and often in a highly  
766 flexible manner.

767

768 Changes in spontaneous activity have previously been shown to contribute to  
769 stimulus responses<sup>50-54</sup> and task-driven changes have been reported in multiple  
770 previous studies<sup>14</sup> but, to our knowledge, have never been given a functional role in  
771 stimulus representation<sup>55</sup>. Here we propose that population-level modulations of  
772 spontaneous activity act as a mechanism supporting the asymmetric representation  
773 of reference and stimuli target in the engaged state. This was clearly the case in  
774 tasks where the passive reference-evoked responses and spontaneous patterns of  
775 activity were not already aligned with respect to the active decoding vector (Fig.8a-d  
776 and Fig8e-h). In those tasks, significant adjustments in spontaneous activity  
777 supported the deployment of a reference/spontaneous space orthogonal to the active  
778 readout-out axis.

779

780 However, this proposed simple linear readout mechanism cannot fully account for the  
781 whole set of responses observed in frontal areas for at least two reasons. First,  
782 projections of reference-elicited activity (in A1) during engagement on an aversive  
783 task still give rise to a non-null, albeit reduced, output contrary to what is observed in  
784 dIFC area recordings. Second, projecting passive data onto the engaged decoding  
785 vector results in symmetric and reduced outputs (data not shown), whereas dIFC  
786 recordings showed on average a complete absence of response during passive state  
787 during the tone-detect task<sup>36</sup>. An additional non-linear gating mechanism likely  
788 operates between primary auditory cortex and frontal areas, further reducing  
789 responses to any stimulus in the passive state and to reference sounds in the active  
790 state. In particular, neurons in higher-order auditory areas could refine the  
791 population-wide, abstracted representation originating in A1 through the proper  
792 combinations of synaptic weights. Such a mechanism could also explain why  
793 individual single units recorded in belt areas of the ferret auditory cortex show a  
794 gradual increase in their selectivity to target stimuli<sup>56</sup>.

### 795 796 **Effects of learning**

797  
798 All the recordings analyzed here were performed on highly trained animals. Several  
799 investigations have reported that training procedures strongly influence neural  
800 representations in primary cortices<sup>57-61</sup>. One may therefore wonder to what extent  
801 our findings, even in the passive state, depend on the prior training history of the  
802 animal<sup>62-64</sup>. To address this question, we examined A1 recordings performed in a  
803 naive ferret exposed to the same stimuli as used in the click-train discrimination task.  
804 Stimulus discrimination was relatively decreased, during both the sound and silent  
805 periods when compared with the decoder accuracy obtained with trained animals  
806 (Fig. S5c,d). In particular, the discrimination performance during the post-stimulus  
807 silence was reduced to chance-levels, while in trained animals it was above chance  
808 even in the passive state. The weak but significant maintained encoding of stimulus  
809 class observed in the passive state with expert ferrets thus appears to be due to the  
810 behavioral training. Discrimination in the passive condition for trained animals also  
811 involved target-specific activity during post-stimulus silence (Fig. 3c,d, bottom  
812 panels), whereas it was not the case for naive ferrets (Fig. S5d), indicating that this  
813 target-driven mechanism is ubiquitously present during the silent period in trained  
814 animals.

815 Interestingly, passive projections of target- and reference-evoked activities  
816 showed variable degrees of asymmetry across tasks (Fig. 3c and 8a,e,i,m). This  
817 observation could be explained by the variability in training duration across ferrets, in  
818 task performance, and in paradigm requirements and complexity. Strikingly, the only  
819 task we examined involving long-term memory (frequency range discrimination task)  
820 exhibited a very strong asymmetry *both* in passive and active states (Fig. 8m). While  
821 asymmetric representation of stimuli was weak in tasks demanding flexible and rapid  
822 attention towards new stimuli (rate discrimination and tone detect tasks), a task  
823 involving long-term memory, such as the frequency range discrimination task, could  
824 engage global reshaping of the neuronal population structure to keep a mnemonic  
825 trace of the behaviorally-relevant stimuli. Interestingly, this target-driven asymmetry in  
826 the passive state came along with a lack of change in the spontaneous population  
827 activity between passive and active state (Fig. 8fo). This observation is in agreement  
828 with the hypothesis that the encoding of stimulus behavioral meaning is mediated by

829 an adjustment of spontaneous population activity, mostly operated in passive state  
830 for this particular task.

831  
832 In summary, we found that task-engagement induces a shift from sensory-  
833 driven to abstract, behavior-driven representations in the primary auditory cortex.  
834 These abstract representations are encoded at a population, but not at a single-  
835 neuron level, and strikingly resemble abstract representations observed in higher-  
836 level cortices. These results suggest that the role of primary sensory cortices is not  
837 limited to encoding sensory features. Instead, primary cortices appear to play an  
838 active role in the task-driven transformation of stimuli into their behavioral meaning  
839 and the translation of that meaning into task-appropriate motor actions.

840  
841  
842  
843

## 844 **Materials and methods**

### 845 **Training and recordings.**

#### 846 *Behavioral training*

847 All experimental procedures conformed to standards specified by the National Institutes of  
848 Health and the University of Maryland Institutional Animal Care and Use Committee  
849 (IACUC). Adult female ferrets, housed in pairs in normal light cycle vivarium, were trained  
850 during the light period on a variety of different behavioral paradigms in a freely moving  
851 training arena. After headpost implantation, the ferrets were retrained while restrained in a  
852 head-fixed holder until they reached performance criterion again. Most of the animals in  
853 these studies were trained on multiple tasks, including the two ferrets trained both on the  
854 click rate discrimination and the tone detect tasks. Three out of four tasks shared the same  
855 basic structure of Go/No-Go avoidance paradigms<sup>65</sup>, in which ferrets were trained in a  
856 conditioned avoidance paradigm to lick water from a spout during the presentation of a class  
857 of reference stimuli and to cease licking after the presentation of a different class of target  
858 stimuli to avoid a mild shock. The positive reinforcement task is detailed below (see *Tone*  
859 *detect task – Aversive conditioning*).

860 Recordings began once the animals had relearned the task in the holder. Each recording  
861 session included epochs of passive sounds presentation without any behavioral response or  
862 reinforcement, followed by an active behavioral epoch where the animals could lick. A post-  
863 passive epoch was then recorded. This sequence of epochs could be repeated multiple  
864 times during a recording session. The table below summarizes the animals and recordings  
865 for each task.

866

<i>Task</i>	<b>Click rate discrimination</b>		<b>Tone detect</b>		<b>Frequency range discrimination</b>
<i>Structure</i>	dIFC	A1		A1	A1
<i>Animals</i>	2 ferrets	2 ferrets		4 ferrets	1 ferret
<i>Conditioning</i>	Aversive	Aversive	Aversive	Appetitive	Aversive
<i>Recorded sessions</i>	- Prepassive - Active - Postpassive	- Prepassive - Active - Postpassive	- Prepassive - Active - Postpassive	- Passive - Active	- Prepassive - Active - Postpassive
<i>Session num.</i>	25 (17 and 8)	18 (9 and 9)	13 (7 and 6)	56 (8,37,2,9)	149
<i>Recorded units</i>	102 (66 and 36)	370 (188 and 182)	202 (129 and 73)	100 (17,72,2,9)	758

867  
868 *Click rate discrimination task.* Two adult female ferrets were trained to discriminate low from  
869 high rate click trains in a Go/No-Go avoidance task. A block of trials consisted of a sequence  
870 of a random number of reference click train trials followed by a target click train trial (except

871 on catch blocks in which 7 reference stimuli were presented with no target). On each trial, the  
872 click train was preceded by a 1.25s neutral noise stimulus (Fig. 1A). Ferrets licked water from  
873 a spout throughout trials containing reference click trains until they heard the target sound.  
874 They learned to stop licking the spout either during the stimulus or after the target click train  
875 ended, in the following 0.4-s time silent response window, in order to avoid a mild shock to  
876 the tongue in a subsequent 0.4 s shock window (Fig.1A). Any lick during this shock window  
877 was punished. The ferrets were first trained while freely-moving daily in a sound-attenuated  
878 test box. Animals were implanted with a headpost when they reached criterion, defined with  
879 a Discrimination Ratio (DR) $\geq$  0.64 where DR = HR \* (1-FA) [Hit Rate, HR=0.8 and False  
880 Alarm, FA=0.2]. They were then retrained head-fixed with the shocks delivered to the tail.  
881 The decision rule was reversed in the 2 animals, as low rates were Go stimuli for one animal  
882 and No-Go for the second one. During each session, rates were kept identical, but were  
883 changed from day to day.

884 *Tone detect task – Aversive conditioning.* The same two ferrets were trained on a tone detect  
885 task previously described<sup>26</sup>. Briefly, a trial consisted of a sequence of 1 to 6 reference white  
886 noise bursts followed by a tonal target (except on catch trials in which 7 reference stimuli  
887 were presented with no target). The frequency of the target pure tone was changed every  
888 day. The animals learned not to lick the spout in a 0.4 s response window starting 0.4 s after  
889 the end of the target. The ferrets were trained until they reached criterion, defined as  
890 consistent performance on the detection task for any tonal target for two sessions with >80%  
891 hit rate accuracy and >80% safe rate for a discrimination rate of >0.65.

892 *Tone detect task – Appetitive conditioning.* 4 ferrets were on an appetitive version of the tone  
893 detect task previously described<sup>30</sup>. On each trial, the number of references presented before  
894 the target varied randomly from one to four. Animals were rewarded with water for licking a  
895 water spout in a response window 0.1–1.0 s after target onset. False alarms were punished  
896 with a timeout when ferrets licked earlier in the trial before the target window. The average  
897 DR during experiments was 0.76. This data set contained sessions with different trial  
898 durations, therefore we analysed separately data from the first 200ms after stimulus onset  
899 and 200ms before stimulus offset. For this task, the passive data was not structured in the  
900 format of successive reference and target trials as in the engaged session but instead the  
901 animal was presented with a block of reference only trials followed by a block of target only  
902 trials separately. This slight change in the structure of the sound presentation did not affect  
903 our results that were highly similar to other tasks but may explain the slightly higher accuracy  
904 of decoding during the initial silence in the passive data. Indeed reference and target trials  
905 were systematically preceded by other reference and target trials, possibly allowing the  
906 decoder to discriminate using remnant activity from the previous trial.

907 *Frequency range discrimination task.* One ferret was trained on a three-frequency-zone  
908 discrimination task with a Go/No-Go paradigm. The three frequency zones were defined  
909 once and for all and the animal had to learn the corresponding frequency boundaries (Low-  
910 Medium: ~500 Hz / Medium-High: ~3400 Hz). Each trial consisted of the presentation of a  
911 single pure tone (0.75-s duration) with a frequency in one of the three zones. A trial began  
912 when the water pump was turned on and the animal licked a spout for water. The ferret  
913 learned to stop licking when it heard a tone falling in the Middle frequency range in order to  
914 avoid punishment (mild shock) but to continue licking if the tone frequency fell in either the  
915 Low or High range. The shock window started 100 ms after tone offset and lasted 400 ms.  
916 The pump was turned off 2 s after the end of the shock window. The learning criterion was  
917 defined as DR>40% in three consecutive sessions of more than 100 trials.

918  
919 *Acoustic stimuli*

920 All sounds were synthesized using a 44 kHz sampling rate, and presented through a free-  
921 field speaker that was equalized to achieve a flat gain. Behavior and stimulus presentation  
922 were controlled by custom software written in Matlab (MathWorks).

923 *Click rate discrimination task.* Target and reference stimuli were preceded by an initial  
924 silence lasting 0.4 s followed by a 1.25 s-long broadband-modulated noise bursts (temporal  
925 orthogonal ripple combinations, TORC<sup>66</sup>) acting as a neutral stimulus, without any behavioral

926 meaning (Fig.1A). Click trains all had the same duration (0.75 s, 0.8 s inter-stimulus interval  
927 of which the last 0.4 s consisted of the response window) and sound level (70 dB SPL).  
928 Rates used were comprised between 6 and 36 Hz (ferret A: references [6 7 8 15] Hz, targets  
929 [24 26 28 30 32 33 36] Hz / ferret L: references [26 28 30 32 36] Hz, targets [6 8 9 16] Hz).

930 *Tone detect task.* Reference sounds were TORC instances. Targets were comprised of pure  
931 tone with frequencies ranging from 125–8000 Hz. Target and reference stimuli were  
932 preceded by an initial silence lasting 0.4 s. Target and reference stimuli all had the same  
933 duration (2 s, 0.8 s inter-stimulus interval whose last 0.4 s consisted of the response window  
934 for the aversive tone detect task) and sound level (70 dB SPL). In the appetitive version of  
935 this paradigm, target and reference duration varied between sessions (0.5–1.0 s, 0.4–0.5-s  
936 interstimulus interval).

937 *Frequency range discrimination task.* The target frequency region was the Medium range  
938 (tone frequencies: 686, 1303 and 2476 Hz) while the reference regions were the Low and  
939 High frequency ranges (100, 190 and 361 Hz; 4705, 8939 and 16884 Hz). Thus the set of  
940 tones included 9 frequencies with 90% increment (~0.9 octave) and spanned a ~7.4 octaves  
941 range. Target and reference stimuli (duration: 0.75 s; level: 70 dB SPL) were preceded by an  
942 initial silence lasting 1.5 s and followed by a 2.4 s silence comprising the shock window (400  
943 ms starting 100 ms after the tone offset).

944

#### 945 *Neurophysiological recordings*

946 To secure stability for electrophysiological recording, a stainless steel headpost was  
947 surgically implanted on the skull (Fritz et al. 2003; Fritz et al. 2010). Experiments were  
948 conducted in a double-walled sound attenuation chamber. Small craniotomies (1–2 mm  
949 diameter) were made over primary auditory cortex prior to recording sessions, each of which  
950 lasted 6–8 h. The A1 and frontal cortex (dorsolateral FC and rostral ASG) regions were  
951 initially located with approximate stereotaxic coordinates and then further identified  
952 physiologically. Recordings were verified as being in A1 according to the presence of  
953 characteristic physiological features (short latency, localized tuning) and to the position of the  
954 neural recording relative to the cortical tonotopic map in A1<sup>67</sup>. Data acquisition was  
955 controlled using the MATLAB software MANTA<sup>68</sup>. Neural activity was recorded using a 24  
956 channel Plexon U-Probe (electrode impedance: ~275 k $\Omega$  at 1 kHz, 75- $\mu$ m inter-electrode  
957 spacing) during the click discrimination task and the aversive version of the tone detect task.  
958 Recordings during the other tasks (frequency range discrimination and appetitive tone detect  
959 task) were done with high-impedance (2-10 M $\Omega$ ) tungsten electrodes (Alpha-Omega and  
960 FHC), using multiple independently moveable electrode drives (Alpha-Omega) to  
961 independently direct up to four electrodes. The electrodes were configured in a square  
962 pattern with ~800  $\mu$ m between electrodes. The probes and the electrodes were inserted  
963 through the dura, orthogonal to the brain's surface, until the majority of channels displayed  
964 spontaneous spiking.

965

#### 966 **Data Analysis**

967 Data analyses were performed in MATLAB (Mathworks, Natick, MA, USA).

#### 968 *Spike sorting*

969 To measure single-unit spiking activity, we digitized and bandpass filtered the continuous  
970 electrophysiological signal between 300 and 6,000 Hz. The tail shock for incorrect responses  
971 introduced a strong electrical artefact and signals recorded during this period were discarded  
972 before processing.

973 Recordings performed with 24 channel Plexotrodes (U-probes) (click discrimination and the  
974 tone detect tasks) were spike sorted using an automatic clustering algorithm (KlustaKwik,<sup>69</sup>),  
975 followed by a manual adjustment of the clusters. Clustering quality was assessed with the  
976 isolation distance, a metrics developed by Harris et al, 2001 which quantifies the increase in  
977 cluster size needed for doubling the number of samples. All clusters showing isolation  
978 distance larger than 20 were considered as single units<sup>70,71</sup>. A total of 82 single units and  
979 288 multi-units were isolated. All analyses were reproduced on both pools of units and

980 qualitatively similar results were obtained (Supplementary Information). We thus combined  
 981 all clusters for the analysis. Spike sorting was performed on merged data sets from pre-  
 982 passive, active and post-passive sessions.

983 For recordings performed with high-impedance tungsten electrodes (frequency range  
 984 discrimination and relative pitch tasks), single units were classified using principal  
 985 components analysis and k-means clustering followed by manual adjustment <sup>26</sup>.  
 986

987 *Depth determination in the click rate discrimination task*

988 Each penetration of the linear electrode array produced a laminar profile of auditory  
 989 responses in A1 across a 1.8 mm depth. Supra- and infragranular layers were determined  
 990 with LFP responses to 100 ms tones recorded during the passive condition. The border  
 991 between superficial and middle-deep layer was defined as the inversion point in correlation  
 992 coefficients between the electrode displaying the shortest response latency and all the other  
 993 electrodes in the same penetration <sup>72,73</sup>.

994  
 995  
 996 *Click reconstruction from neural data*

997 Optimal prior reconstruction method <sup>74</sup> was used to reconstruct stimulus waveform from click-  
 998 elicited neural activity. Units with spontaneous firing rate larger than 2 spikes/s in at least one  
 999 condition were considered for this analysis. Neuronal activity was binned at 10 ms in time

1000 with a 1-ms time step. For each trial, we defined  $S^k(t)$   $S^k(t)$  the stimulus waveform of trial  $k$   
 1001 ( $t \in [1, T]$ ) and  $r_i^k(t)$   $r_i^k(t)$  the binned firing rate of each neuron  $i \in [1, N]$  where  $t \in [1, T + \tau]$  with  
 1002  $\tau$  the considered delay in the neuronal response. A linear mapping was assumed between  
 1003 the neuronal responses and the stimulus:  
 1004

$$1005 \quad S^k(t) = \sum_{i=1}^N \sum_{\delta=0}^{\tau} g_i(\delta) r_i^k(t + \delta) S^k(t) = \sum_{i=1}^N \sum_{\delta=0}^{\tau} g_i(\delta) r_i^k(t + \delta) \quad (1)$$

1006

1007 for unknown coefficients  $g_i(\delta)$ . Equation (1) was rewritten as:

$$1008 \quad S^k = GR^k S^k = GR^k \quad (2)$$

1009  
 1010

$$1011 \quad \text{with} \quad R^k = \begin{pmatrix} R_1^k \\ R_2^k \\ \vdots \\ R_N^k \end{pmatrix} \quad R^k = \begin{pmatrix} R_1^k \\ R_2^k \\ \vdots \\ R_N^k \end{pmatrix} \quad \text{and} \quad R_i^k = \begin{pmatrix} r_i^k(0) & r_i^k(1) & \dots & r_i^k(T) \\ r_i^k(1) & r_i^k(2) & \dots & r_i^k(T+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_i^k(\tau) & r_i^k(1+\tau) & \dots & r_i^k(T+\tau) \end{pmatrix}$$

$$1012 \quad R_i^k = \begin{pmatrix} r_i^k(0) & r_i^k(1) & \dots & r_i^k(T) \\ r_i^k(1) & r_i^k(2) & \dots & r_i^k(T+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_i^k(\tau) & r_i^k(1+\tau) & \dots & r_i^k(T+\tau) \end{pmatrix} \text{ the lagged neuronal}$$

1013

1014 response,  $G = (G_1, G_2 \dots G_N)$   $G = (G_1 \ G_2 \ \dots \ G_N)$  and  $G_i = (g_i(0), g_i(1) \dots g_i(\tau))$

1015  $G_i = (g_i(0) \ g_i(1) \ \dots \ g_i(\tau))$  the corresponding reconstruction filter. The estimate  $\hat{G}$  is  
 1016 produced by least-square fitting  
 1017

$$1018 \quad \hat{G} = S \left( \sum_{k=1}^K (R^k)^t \right) \left( \sum_{k=1}^K (R^k)^t R^k \right)^{-1} \hat{G} = S \left( \sum_{k=1}^K (R^k)^t \right) \left( \sum_{k=1}^K (R^k)^t R^k \right)^{-1} \quad (3)$$

1019

1020 Before the inversion in the previous formula, a single value decomposition was used to  
1021 eliminate the noisy components of the auto-correlation matrix. The maximal number of  
1022 components retained was empirically set to 70. Once the values  $\hat{G}$   $\hat{G}$  were fitted on all the  
1023 trials but one, the reconstructed stimulus  $\hat{S}^k$   $\hat{S}^k$  was defined as  $\hat{S}^k = \hat{G}_R^k \hat{S}^k = \hat{G}R^k$  with the  
1024 neuronal response  $R$  of the remaining run. Each trial was left out in turn. Reconstruction error  
1025 was quantified with the mean-squared error (MSE) of the reconstructed stimulus. One  
1026 passive and active reconstruction filters were fitted for each type of stimulus (reference and  
1027 target) in every session.

#### 1028 *Modulation index*

1029 To evaluate changes in a given parameter  $X$  (firing rate, vector strength) at the level of the  
1030 individual unit, we define the modulation index to compare situation 1 and 2 as for each  
1031 neuron as:  
1032

$$1033 \quad MI = \frac{X_1 - X_2}{X_1 + X_2}$$

1034

1035 As a measure of the enhancement of target projection relative to reference projection in the  
1036 task engaged state we used the following index (referred to target enhancement index in the  
text)

$$MI = (d(Targ_{eng}) - d(Targ_{pass})) - (d(Ref_{eng}) - d(Ref_{pass}))$$

1037 where  $d$  is the distance from baseline.

1038 When simply measuring the asymmetry between reference and target in condition  $X$ , we  
1039 used the following index (Fig. 5b; 7d,h,l,p; S9c,f,i):

$$Index = d(Targ_X) - d(Ref_X)$$

#### 1040 *Vector strength*

1041 Vector strength (VS) allows to measure how tightly spiking activity is locked to one phase of  
1042 a stimulus. If all spikes at exactly the same phase, VS is one whereas if firing is uniformly  
1043 distributed over phases VS is 0. It is defined in Goldberg & Brown 1969 as  
1044

$$1045 \quad VS = \frac{\sqrt{(\sum_{i=1}^n \cos \theta_i)^2 + (\sum_{i=1}^n \sin \theta_i)^2}}{n} \text{ where } \theta_i \text{ is the phase of spike } i$$

1046 Significance was assessed using Rayleigh's statistic,  $p = e^{-nr^2}$ , where  $r$  is the vector strength  
1047 and used  $p < 0.001$  as the criterion for significant phase locking consistent with previous  
1048 work<sup>76</sup>.

#### 1049 *Linear discriminant classifier performance*

1050 To evaluate the accuracy with which single-trial population responses could be classified  
1051 according to the presented stimulus (reference or target), we trained and tested a linear  
1052 discriminant classifier<sup>39,77</sup> using cross validation (FigS3).

1053 Trial by trial pseudo-population firing rate vectors were constructed for each 100ms time bin  
1054 using units from all sessions and both animals. Training and testing sets were constructed by  
1055 randomly selecting equal numbers (15) of reference and target trials for each unit. All  
1056 contribution of noise correlations among neurons are therefore destroyed by this procedure  
1057 as the pseudo-population vector contains activity of units recorded on different days and on  
1058

1059 different trials. Since correlations between neurons can affect population coding<sup>33</sup> and are  
 1060 modified by task engagement<sup>13</sup>,  
 1061 The classifier was trained for each time bin using the average pseudo-population vectors  $c_{R,t}$   
 1062 and  $c_{T,t}$  calculated from a random selection of an equal number of reference and target trials.  
 1063 These vectors define at time bin  $t$  the decoding vector  $w_t$  given by

$$1064 \quad w_t = c_{T,t} - c_{R,t}$$

1065 and the bias  $b_t$  given by

$$1066 \quad b_t = \frac{-(c_{R,t} \times w_t + c_{T,t} \times w_t)}{2}$$

1067 we also used Fisher discriminant analysis in which the decoding vector is defined as :

$$1068 \quad w_t = Cov^{-1}(c_{T,t} - c_{R,t})$$

1069 where Cov is the covariance matrix, which allows to correct the decoding vector by taking into account the trial  
 1070 by trial correlations between units

1071 These define the decision rule for a new population vector  $x$ ,

$$y(x) = w_t^T \times x + b_t$$

1072  $y(x) > 0$ ,  $x$  is classified as a target

$y(x) < 0$ ,  $x$  is classified as a reference

1073 This rule was applied to an equal number of reference and target testing trials drawn from  
 1074 the remaining trials that were not used to train the classifier. The proportion of correctly  
 1075 classified trials gave the accuracy of the classifier. Cross-validation was performed 400 times  
 1076 by randomly picking training and testing data to estimate the average and variance of  
 1077 accuracy. This allowed comparing the performance of classification in two behavioral states  
 1078 by constructing confidence intervals from the cross-validation. Note that this limits p-value  
 1079 estimate to a minimum of  $1/400=0.0025$ .

1080

#### 1081 *Random performance*

1082 To evaluate whether the classifier performance is higher than chance, the classifier was  
 1083 trained and tested on surrogate data sets constructed by shuffling the labels ('reference' and  
 1084 'target') of trials. For each of 100 label permutations, cross-validation was performed 100  
 1085 times. This allows comparing the performance of classification with chance levels by  
 1086 constructing confidence intervals from the cross-validation and from the random shuffled  
 1087 permutations.

1088

1089

#### 1090 *Classifier evolution*

1091 When studying the evolution of population encoding (Fig. S6), we defined early sound, late  
 1092 sound, and silence periods as 1700-1900 ms, 2200-2400 ms and 2700-2900 ms (equal  
 1093 duration for comparison) relative to trial onset. The classifier was trained on randomly chosen  
 1094 trials from one time period and then tested on trials at all other 100ms time bins.

1095 We also constructed matrices showing the accuracy of the classifier trained and tested at all  
 1096 100ms time bins and evaluated whether these values are higher than chance using  
 1097 surrogate data sets by shuffling labels as described above.

1098 When comparing the classifier during sound and silence periods across tasks (Fig. 7), the  
 1099 following periods were used:

1100

	Click rate discrimination	Aversive tone detect	Appetitive tone detect	Frequency range discrimination
Sound	1.6-2s	0.4-0.8s	0-0.1 after stim onset	1.5-1.9s

Silence	2.5-2.9s	2.5-2.9s	0-0.1 after stim offset	2.4-2.8s
---------	----------	----------	-------------------------	----------

1101

### 1102 *Projection onto decoding vectors*

1103 To study the contribution of reference and target trials to classifier performance, we projected  
1104 population firing vectors at each time bin onto decoding vectors calculated during the sound  
1105 and silence periods as defined above. Before projection, the mean spontaneous activity of  
1106 each unit was subtracted from its firing rate throughout the whole trial. Deviations from 0 of  
1107 the projection show activity deviating from spontaneous activity along the decoding axis.

### 1108 *Controlling for lick-responsive neurons*

1109 In order to control for the contribution of units directly linked with task-related motor activity to  
1110 our results, we combined reconstruction and decoding methods to identify and remove lick-  
1111 responsive neurons so that linear classification no longer yielded any licking-related  
1112 information. The approach comprised the following steps:

- 1113 - Optimal prior reconstruction (described in *Click reconstruction from neural data*) was  
1114 used to reconstruct lick-activity separately for each unit.
  - 1115 - Reconstruction values for each unit were then sampled at the time of licks and at  
1116 randomly selected times without licking. These values were used to construct  
1117 population vectors of lick and non-lick activity.
  - 1118 - A linear classifier (described in *Linear discriminant classifier performance*) was  
1119 trained and tested using cross-validation to distinguish lick from non-lick events.
  - 1120 - Reconstruction values and classification was also performed on random data  
1121 obtained by reconstructing the licking activity of a session with the neural activity of a  
1122 subsequent session. This made it possible to establish the distribution of accuracy for  
1123 randomized data.
  - 1124 - The accuracy of classification was compared between the true data and the  
1125 randomized data sets and a p-value was calculated by counting the number of  
1126 permutations showing better accuracy for the randomized data than the true data.
  - 1127 - We progressively removed units, starting with those with highest classifier weights,  
1128 which reduced the accuracy of classification, until the p-value of population  
1129 classification rose above 0.4. This indicated that the remaining units contained no  
1130 more information about lick events than randomized data.
  - 1131 - Only the units remaining after this procedure were used to re-analyze the data and  
1132 verify that reliable classification and difference in projections of reference and tone  
1133 trials did not rely on the difference in licking activity between the two trials.
- 1134 For the click rate discrimination task only a subset of sessions (15/18) had reliable  
1135 recordings of all lick events, so the analysis was done on 308 units (not 370), 277  
1136 units were identified as non-lick related. For the appetitive tone task 99/100 units, for  
1137 the aversive tone task 161/202 and for the frequency range discrimination 520/758.

1138

### 1139 *Gaussian-process factor analysis*

1140 To visualize neural trajectories of the large population of units recorded in A1, we used  
1141 Gaussian-process factor analysis as described in <sup>78</sup>. This method has the advantage over  
1142 more traditional methods of dimensionality reduction such as PCA of jointly performing both  
1143 the binning/smoothing steps and the dimensionality reduction.

1144

### 1145 *Statistics*

1146 Statistics on classifier performance relied on p-value estimation using cross-validation. For  
1147 each statistical analysis provided in the manuscript, the Kolmogorov–Smirnov normality test  
1148 was first performed on the data. As the data failed to meet the normality criterion, statistics  
1149 relied on non-parametric tests. When performing systematic multiple tests, the Bonferroni  
1150 correction was applied.

1151

### 1152 *Data availability*

1153 The data that support the findings of this study are available from the corresponding author  
1154 upon reasonable request.

1155

1156 *Code availability*

1157 Code used in the article can be supplied upon request by writing to the corresponding author.

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187

1188

1189

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204 **REFERENCES**

- 1205 1. Chechik, G. *et al.* Reduction of information redundancy in the ascending  
1206 auditory pathway. *Neuron* **51**, 359–368 (2006).
- 1207 2. Chechik, G. & Nelken, I. Auditory abstraction from spectro-temporal features to  
1208 coding auditory entities. *Proc Natl Acad Sci U S A* **109**, 18968–18973 (2012).
- 1209 3. de Lafuente, V. & Romo, R. Neural correlate of subjective sensory experience  
1210 gradually builds up across cortical areas. *Proc Natl Acad Sci U S A* **103**,  
1211 14266–14271 (2006).
- 1212 4. Siegel, M., Buschman, T. J. & Miller, E. K. BRAIN PROCESSING. Cortical  
1213 information flow during flexible sensorimotor decisions. *Science* (80-. ). **348**,  
1214 1352–1355 (2015).
- 1215 5. Vergara, J., Rivera, N., Rossi-Pool, R. & Romo, R. A Neural Parametric Code  
1216 for Storing Information of More than One Sensory Modality in Working Memory.  
1217 *Neuron* **89**, 54–62 (2016).
- 1218 6. D’Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory.  
1219 *Annu. Rev. Psychol.* **66**, 115–42 (2015).
- 1220 7. de Lafuente, V. & Romo, R. Neuronal correlates of subjective sensory  
1221 experience. *Nat Neurosci* **8**, 1698–1703 (2005).
- 1222 8. Lemus, L., Hernández, A. & Romo, R. Neural codes for perceptual  
1223 discrimination of acoustic flutter in the primate auditory cortex. *Proc Natl Acad*  
1224 *Sci U S A* **106**, 9471–9476 (2009).
- 1225 9. Yildiz, I. B., Mesgarani, N. & Deneve, S. Predictive Ensemble Decoding of  
1226 Acoustical Features Explains Context-Dependent Receptive Fields. *J.*  
1227 *Neurosci.* **36**, 12338–12350 (2016).
- 1228 10. Sloas, D. C. *et al.* Interactions across Multiple Stimulus Dimensions in Primary  
1229 Auditory Cortex. *eNeuro* **3**, 1–7 (2016).
- 1230 11. Bizley, J. K., Walker, K. M. M., Nodal, F. R., King, A. J. & Schnupp, J. W. H.  
1231 Auditory cortex represents both pitch judgments and the corresponding  
1232 acoustic cues. *Curr Biol* **23**, 620–625 (2013).
- 1233 12. Niwa, M., Johnson, J. S., O’Connor, K. N. & Sutter, M. L. Active engagement  
1234 improves primary auditory cortical neurons’ ability to discriminate temporal  
1235 modulation. *J Neurosci* **32**, 9323–9334 (2012).
- 1236 13. Downer, J. D., Niwa, M. & Sutter, M. L. Task engagement selectively  
1237 modulates neural correlations in primary auditory cortex. *J Neurosci* **35**, 7565–  
1238 74 (2015).
- 1239 14. Otazu, G. H., Tai, L.-H., Yang, Y. & Zador, A. M. Engaging in an auditory task  
1240 suppresses responses in auditory cortex. *Nat Neurosci* **12**, 646–654 (2009).
- 1241 15. Brosch, M. Nonauditory Events of a Behavioral Procedure Activate Auditory  
1242 Cortex of Highly Trained Monkeys. *J. Neurosci.* **25**, 6797–6806 (2005).
- 1243 16. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral  
1244 State in Mouse Visual Cortex. *Neuron* **65**, 472–479 (2010).
- 1245 17. Schneider, D. M., Nelson, A. & Mooney, R. A synaptic and circuit basis for  
1246 corollary discharge in the auditory cortex. *Nature* **513**, 189–194 (2014).
- 1247 18. Zhou, M. *et al.* Scaling down of balanced excitation and inhibition by active  
1248 behavioral states in auditory cortex. *Nat. Neurosci.* **17**, 841–50 (2014).
- 1249 19. Rodgers, C. C. & DeWeese, M. R. Neural correlates of task switching in  
1250 prefrontal cortex and primary auditory cortex in a novel stimulus selection task  
1251 for rodents. *Neuron* **82**, 1157–1170 (2014).
- 1252 20. Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y. & Petersen,  
1253 C. C. H. Membrane potential correlates of sensory perception in mouse barrel

- 1254 cortex. *Nat Neurosci* **16**, 1671–1677 (2013).
- 1255 21. Shuler, M. G. & Bear, M. F. Reward Timing in the Primary Visual Cortex.  
1256 *Science* (80-. ). **311**, 1606–1610 (2006).
- 1257 22. Petreanu, L. *et al.* Activity in motor-sensory projections reveals distributed  
1258 coding in somatosensation. *Nature* **489**, 299–303 (2012).
- 1259 23. Ohl, F. W., Scheich, H. & Freeman, W. J. Change in pattern of ongoing cortical  
1260 activity with auditory category learning. *Nature* **412**, 733–736 (2001).
- 1261 24. Quirk, G. J., Armony, J. L. & LeDoux, J. E. Fear conditioning enhances  
1262 different temporal components of tone-evoked spike trains in auditory cortex  
1263 and lateral amygdala. *Neuron* **19**, 613–624 (1997).
- 1264 25. Kuchibhotla, K. V *et al.* Parallel processing by cortical inhibition enables  
1265 context-dependent behavior. *Nat. Neurosci.* 1–14 (2016). doi:10.1038/nn.4436
- 1266 26. Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of  
1267 spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **6**,  
1268 1216–1223 (2003).
- 1269 27. Fritz, J. B., Elhilali, M. & Shamma, S. A. Differential dynamic plasticity of A1  
1270 receptive fields during multiple spectral tasks. *J. Neurosci.* **25**, 7623–35 (2005).
- 1271 28. Fritz, J. B., Elhilali, M. & Shamma, S. a. Adaptive changes in cortical receptive  
1272 fields induced by attention to complex sounds. *J Neurophysiol* **98**, 2337–2346  
1273 (2007).
- 1274 29. Atiani, S., Elhilali, M., David, S. V, Fritz, J. B. & Shamma, S. A. Task Difficulty  
1275 and Performance Induce Diverse Adaptive Patterns in Gain and Shape of  
1276 Primary Auditory Cortical Receptive Fields. *Neuron* **61**, 467–480 (2009).
- 1277 30. David, S. V, Fritz, J. B. & Shamma, S. A. Task reward structure shapes rapid  
1278 receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A* **109**,  
1279 2144–2149 (2012).
- 1280 31. Yin, P., Johnson, J. S. & Sutter, M. L. Coding of Amplitude Modulation in  
1281 Primary Auditory Cortex. *J. Neurophysiol.* 582–600 (2010).  
1282 doi:10.1152/jn.00621.2010
- 1283 32. Averbeck, B. & Lee, D. Effects of noise correlations on information encoding  
1284 and decoding. *J. Neurophysiol.* 3633–3644 (2006).  
1285 doi:10.1152/jn.00919.2005.Effects
- 1286 33. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population  
1287 coding and computation. *Nat Rev Neurosci* **7**, 358–366 (2006).
- 1288 34. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by  
1289 reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
- 1290 35. Downer, J. D., Rapone, B., Verhein, J., O'Connor, K. N. & Sutter, M. L.  
1291 Feature-Selective Attention Adaptively Shifts Noise Correlations in Primary  
1292 Auditory Cortex. *J. Neurosci.* **37**, 5378–5392 (2017).
- 1293 36. Fritz, J. B., David, S. V, Radtke-Schuller, S., Yin, P. & Shamma, S. A. Adaptive,  
1294 behaviorally gated, persistent encoding of task-relevant auditory information in  
1295 ferret frontal cortex. *Nat. Neurosci.* **13**, 1011–1019 (2010).
- 1296 37. Ahveninen, J. *et al.* Attention-driven auditory cortex short-term plasticity helps  
1297 segregate relevant sounds from noise. *Proc Natl Acad Sci U S A* **108**, 4182–  
1298 4187 (2011).
- 1299 38. Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Activity related to  
1300 perceptual judgment and action in primary auditory cortex. *J Neurosci* **32**,  
1301 3193–3210 (2012).
- 1302 39. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T.  
1303 Dynamic population coding of category information in inferior temporal and

- 1304 prefrontal cortex. *J. Neurophysiol.* **100**, 1407–1419 (2008).
- 1305 40. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks.
- 1306 *Nature* **497**, 585–590 (2013).
- 1307 41. Saez, A., Rigotti, M., Ostojic, S., Fusi, S. & Salzman, C. D. Abstract Context
- 1308 Representations in Primate Amygdala and Prefrontal Cortex. *Neuron* **87**, 869–
- 1309 881 (2015).
- 1310 42. Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in
- 1311 premotor cortex during motor planning. *Nature* **532**, 459–64 (2016).
- 1312 43. Letzkus, J. J. *et al.* A disinhibitory microcircuit for associative fear learning in
- 1313 the auditory cortex. *Nature* **480**, 331–335 (2011).
- 1314 44. Parikh, V., Kozak, R., Martinez, V. & Sarter, M. Prefrontal Acetylcholine
- 1315 Release Controls Cue Detection on Multiple Timescales. *Neuron* **56**, 141–154
- 1316 (2007).
- 1317 45. Sheikhattar, A., Miran, S., Fritz, J. B., Shamma, S. A. & Babadi, B. Probing the
- 1318 Functional Circuitry Underlying Auditory Attention via Dynamic Granger
- 1319 Causality Analysis. *Proc. 50th Asilomar Conf. Signals, Syst. Comput.* (2016).
- 1320 46. Budinger, E., Laszcz, A., Lison, H., Scheich, H. & Ohl, F. W. Non-sensory
- 1321 cortical and subcortical connections of the primary auditory cortex in Mongolian
- 1322 gerbils: Bottom-up and top-down processing of neuronal information via field
- 1323 A1. *Brain Res.* **1220**, 2–32 (2008).
- 1324 47. Winkowski, D. E. *et al.* Orbitofrontal Cortex Neurons Respond to Sound and
- 1325 Activate Primary Auditory Cortex Neurons. *Cereb. Cortex* 1–12 (2017).
- 1326 doi:10.1093/cercor/bhw409
- 1327 48. Bizley, J. K., Bajo, V. M., Nodal, F. R. & King, A. J. Cortico-cortical connectivity
- 1328 within ferret auditory cortex. *J. Comp. Neurol.* **2210**, 2187–2210 (2015).
- 1329 49. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity
- 1330 in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**,
- 1331 440–8 (2014).
- 1332 50. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity:
- 1333 explanation of the large variability in evoked cortical responses. *Science (80-. )*.
- 1334 **273**, 1868–1871 (1996).
- 1335 51. Luczak, A., Bartho, P. & Harris, K. D. Gating of Sensory Input by Spontaneous
- 1336 Cortical Activity. *J. Neurosci.* **33**, 1684–1695 (2013).
- 1337 52. Tatti, R. & Maffei, A. Synaptic dynamics: How network activity affects neuron
- 1338 communication. *Curr. Biol.* **25**, R278–R280 (2015).
- 1339 53. Harris, K. D. & Thiele, A. Cortical state and attention. *Nat Rev Neurosci* **12**,
- 1340 509–523 (2011).
- 1341 54. Carcea, A. I., Insanally, M. N. & Froemke, R. C. Dynamics of cortical activity
- 1342 during behavioral engagement and auditory perception. *Nat. Commun.* **8**, 1–12
- 1343 (2017).
- 1344 55. Driver, J. & Frith, C. Shifting baselines in attention research. *Nat. Rev.*
- 1345 *Neurosci.* **1**, 147–148 (2000).
- 1346 56. Atiani, S. *et al.* Emergent selectivity for task-relevant stimuli in higher-order
- 1347 auditory cortex. *Neuron* **82**, 486–499 (2014).
- 1348 57. Makino, H. & Komiyama, T. Learning enhances the relative impact of top-down
- 1349 processing in the visual cortex. *Nat. Neurosci.* **18**, 1116–1122 (2016).
- 1350 58. Chen, J. L. *et al.* Pathway-specific reorganization of projection neurons in
- 1351 somatosensory cortex during learning. *Nat. Neurosci.* **18**, 1101–1108 (2015).
- 1352 59. Kato, H. K., Gillet, S. N. & Isaacson, J. S. Flexible Sensory Representations in
- 1353 Auditory Cortex Driven by Behavioral Relevance. *Neuron* **88**, 1027–1039

- 1354 (2015).
- 1355 60. Poort, J. *et al.* Learning Enhances Sensory and Multiple Non-sensory  
1356 Representations in Primary Visual Cortex. *Neuron* **86**, 1478–1490 (2015).
- 1357 61. Weinberger, N. M., Javid, R. & Lapan, B. Long-term retention of learning-  
1358 induced receptive-field plasticity in the auditory cortex. *Proc. Natl. Acad. Sci. U.*  
1359 *S. A.* **90**, 2394–2398 (1993).
- 1360 62. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics  
1361 by sensory input during natural vision. *Nature* **431**, 573–578 (2004).
- 1362 63. Fiser, J., Berkes, P., Orban, G. & Lengyel, M. Statistically optimal perception  
1363 and learning: from behavior to neural representations. *Trends Cogn. Sci.*  
1364 (2010). doi:10.1016/j.tics.2010.0
- 1365 64. Berkes, P., Orban, G., Lengyel, M. & Fiser, J. Spontaneous Cortical Activity  
1366 Reveals Hallmarks of an Optimal Internal Model of the Environment. *Science*  
1367 (80- ). **331**, 83–87 (2011).
- 1368 65. Heffner, H. E. & Heffner, R. S. Conditioned Avoidance. *Methods Comp.*  
1369 *Psychoacoustics* 79–93 (1995).
- 1370 66. Klein, D. J., Depireux, D. A., Simon, J. Z. & Shamma, S. A. Robust  
1371 Spectrotemporal and Reverse Correlation and for the Auditory and System:  
1372 and Optimizing Stimulus and Design. *J. Comput. Neurosci.* **9**, 85–111 (2000).
- 1373 67. Shamma, S. a, Fleshman, J. W., Wiser, P. R. & Versnel, H. Organization of  
1374 response areas in ferret primary auditory cortex. *J. Neurophysiol.* **69**, 367–383  
1375 (1993).
- 1376 68. Englitz, B., David, S. V, Sorenson, M. D. & Shamma, S. A. MANTA-an open-  
1377 source, high density electrophysiology recording suite for MATLAB. *Front*  
1378 *Neural Circuits* **7**, 69 (2013).
- 1379 69. Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. & Buzsáki, G. Accuracy of  
1380 tetrode spike separation as determined by simultaneous intracellular and  
1381 extracellular measurements. *J. Neurophysiol.* **84**, 401–414 (2000).
- 1382 70. Belliveau, L. A. C., Lyamzin, D. R. & Lesica, N. A. The neural representation of  
1383 interaural time differences in gerbils is transformed from midbrain to cortex. *J.*  
1384 *Neurosci.* **34**, 16796–808 (2014).
- 1385 71. Garcia-Lazaro, J. A., Shepard, K. N., Miranda, J. A., Liu, R. C. & Lesica, N. A.  
1386 An overrepresentation of high frequencies in the mouse inferior colliculus  
1387 supports the processing of ultrasonic vocalizations. *PLoS One* **10**, (2015).
- 1388 72. Kajikawa, Y. & Schroeder, C. E. How local is the local field potential? *Neuron*  
1389 **72**, 847–858 (2011).
- 1390 73. Linden, J. F. & Schreiner, C. E. Columnar transformations in auditory cortex? A  
1391 comparison to visual and somatosensory cortices. *Cereb. Cortex* **13**, 83–89  
1392 (2003).
- 1393 74. Mesgarani, N., David, S. V, Fritz, J. B. & Shamma, S. A. Influence of Context  
1394 and Behavior on Stimulus Reconstruction From Neural Activity in Primary  
1395 Auditory Cortex. *J. Neurophysiol.* **102**, 3329–3339 (2009).
- 1396 75. Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior  
1397 olivary complex to dichotic tonal stimuli: some physiological mechanisms of  
1398 sound localization. *J. Neurophysiol.* **32**, 613–636 (1969).
- 1399 76. Gao, X. & Wehr, M. A Coding Transformation for Temporally Structured  
1400 Sounds within Auditory Cortical Neurons. *Neuron* **86**, 292–303 (2015).
- 1401 77. Bishop, C. M. *Pattern Recognition and Machine Learning. Pattern Recognition*  
1402 **4**, (2006).
- 1403 78. Yu, B. M. *et al.* Gaussian-Process Factor Analysis for Low-Dimensional Single-

1404 Trial Analysis of Neural Population Activity. *J. Neurophysiol.* **102**, 614–635  
1405 (2009).  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452

1453 **Supplementary information**  
 1454 **Comparison of results in single and multiunits**

1455 All analyses in the main section of the paper concerning the click train discrimination  
 1456 task combine results from single units (isolation distance > 20, see Methods) and  
 1457 multi-units because we found no differences concerning their general properties (see  
 1458 table 1) and the main population-level results of the paper (see table 2) were  
 1459 maintained using SU activity only, although the power of the analysis was of course  
 1460 reduced.

	SU	MU	Comparison
MI : baseline	0.14 +/- 0.03 (***)	0.19 +/- 0.02 (***)	p= 0.22
MI : evoked	0.04 +/- 0.05 (ns)	- 0.05 +/- 0.06 (ns)	p=0.22
MI : vector strength	0.05 +/- 0.006 (***)	0.04 +/- 0.0075 (***)	p=0.25
Ref FR pass. – Snd	7.45 +/- 0.70	6.67 +/- 0.88	p=0.48
Ref FR eng. – Snd	9.14 +/-0.86	8.38 +/- 1.06	p=0.57
Targ FR pass. – Snd	7.78 +/- 0.72	6.15 +/- 0.77	p=0.12
Tar FR eng. – Snd	9.9 +/- 0.93	7.9 +/- 0.97	p=0.15
Ref FR pass. – Sil	6.34 +/-0.64	5.3 +/- 0.68	p=0.25
Ref FR eng. – Sil	7.96 +/-0.76	7.65 +/- 0.94	p=0.79
Targ FR pass. – Sil	6.31 +/-0.67	5.4 +/- 0.76	p=0.36
Targ FR eng – Sil.	8.56 +/-0.84	7.26 +/- 0.99	p=0.32

1461 **Table 1** Comparison of unit properties for single and multi units. Mean +/- s.e.m are  
 1462 given for each value and the comparison between SU and MU is performed using a  
 1463 ttest. For modulation indexes (first three lines), the significance compared to zero is  
 1464 given in brackets. These results are identical to those found in the main paper.  
 1465

1466 To verify that the population-level results were maintained SU data, despite the  
 1467 reduced number of units (82 SU units, 370 total units used in main paper), we  
 1468 recapitulate below the main results using SU activity alone.  
 1469

	Mean [C.I.] – signif. of comparison
Sound accuracy pass. and eng.	0.97 [0.95:0.99] - 0.98 [0.94:1] <b>NS</b>
Silence accuracy pass. and eng.	0.59 [0.52:0.66] - 0.78 [0.69:0.87] *
Sound: ref and target projected values pass.	29 [25:36] - 26 [18:33] <b>NS</b>
Silence: ref and target projected values pass.	6 [4:8] - 12 [6:16] <b>NS</b>
Sound: ref and target projected values eng.	16 [8:23] - 44 [33:55] **
Silence: ref and target projected values eng.	2 [0.7:4] - 37 [33:42] **

1470 **Table 2** Recapitulation of important results using SU activity alone.  
 1471

1472 We found that the significant increase in accuracy during the silence with task  
1473 engagement was maintained after restriction to SU activity. We also observed the  
1474 significantly greater role played by target evoked activity in the engaged state after  
1475 projection (as in Fig. 3) using SU activity alone ( $p < 0.0025$ ). The only difference with  
1476 results given in the main paper is that in the passive state during the silence the  
1477 stronger contribution of target activity did not achieve significance as in Fig. 3d,  
1478 bottom.

1479

### 1480 **Population-encoding dynamics change between conditions**

1481 In the analyses reported in the main text, we trained a classifier at each time point in  
1482 the trial, and used it to evaluate stimulus discrimination at the same time point in  
1483 held-out trials. To assess how much the underlying encoding changes over the trial,  
1484 we used two procedures. First, we directly compared the classifiers determined at  
1485 different time-bins by computing the correlation between them (Fig. S6a,c). Second,  
1486 we used the classifier obtained at three different trial epochs (early and late stimulus,  
1487 post-stimulus silence) to classify the neural activity along the whole trials (Fig. S6b,d).  
1488 If the encoding of stimulus underlying stimulus discrimination changes over time in  
1489 the trial, a classifier trained on one time point will lead to a lower discrimination  
1490 performance at other times.

1491 In the passive condition, we found that changes in encoding over time are weak. The  
1492 encoding was highly homogeneous within stimulus presentation and during the post-  
1493 sound silence (Fig. S6a). Consistent with this view, classifiers trained during the early  
1494 or the late phases of the stimulus presentation could be used efficiently at all other  
1495 times during stimulus presentation without an appreciable drop in accuracy (Fig. S6b,  
1496 brown and orange curves). In contrast, the same classifier led to chance-level  
1497 discrimination at time points after stimulus presentation. Conversely a classifier  
1498 trained after stimulus presentation led to chance-level performance during stimulus  
1499 presentation (Fig. S6b, yellow curve). In the passive condition, the neural encoding  
1500 that underlies stimulus discrimination therefore appears to change very little during  
1501 stimulus presentation, and shifts abruptly afterwards.

1502 A different picture emerged when animals were engaged in the task. The encoding  
1503 appeared to change more progressively over the trial (Fig. S6c), and a classifier  
1504 trained at one point systematically led to reduced discrimination performance at other  
1505 time points (Fig. S6d). Moreover, no sharp transition was apparent at the time the  
1506 stimulus was switched off. In particular, a classifier trained during the stimulus  
1507 presentation led to a significant discrimination performance after stimulus  
1508 presentation (Fig. S6d, brown and orange curves). Conversely, a classifier  
1509 determined during the post-sound silence led to an above chance and progressively  
1510 increasing discrimination performance during stimulus presentation (Fig. S6d, yellow  
1511 curve).

1512 Altogether, in the engaged condition, the population encoding underlying stimulus  
1513 discrimination therefore appeared to progressively shift from a representation purely  
1514 along a stimulus-driven axis, where categorical information was present but  
1515 uncorrelated with behavior (Fig. 3c top panel), to a representation along a decision-  
1516 related axis, which was directly correlated with the behavioral action (Fig. 3e bottom  
1517 panel).

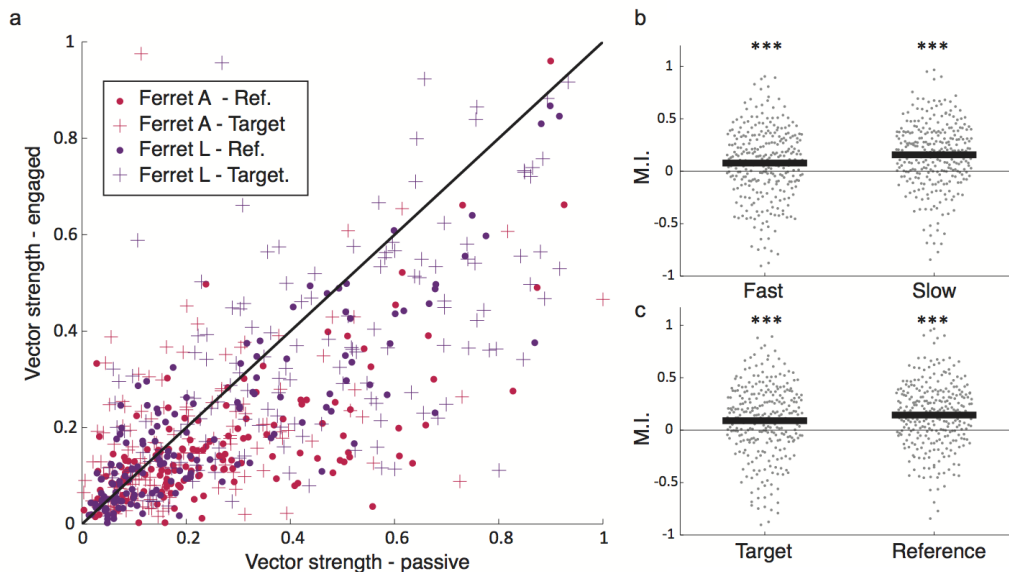
1518

1519

1520

1521

1522 **Supplementary Figures**  
1523



**FigS1.** Changes in stimulus entrainment between passive and engaged conditions

1524

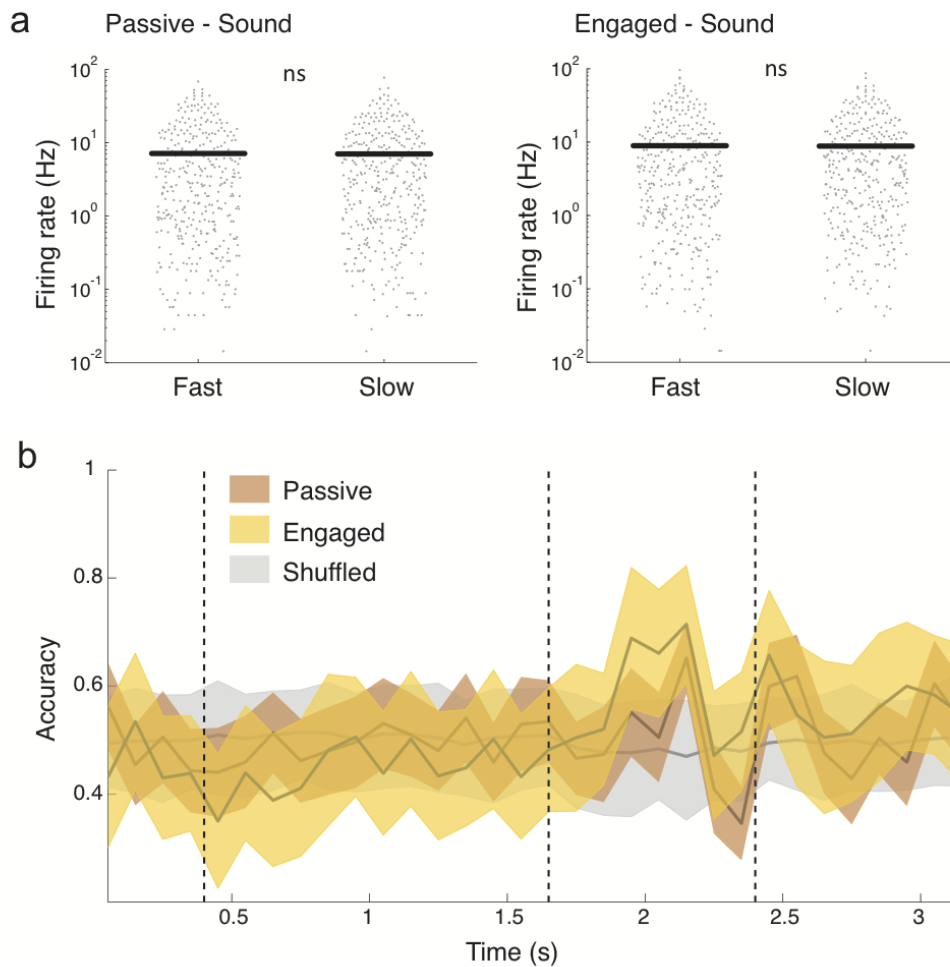
**Fig S1**

a. For each unit the vector strength for the reference and target click train is plotted in the engaged state vs the passive state. Animals are given in different colours and stimuli as different markers. Note that most points are below the  $x=y$  line, showing higher phase locking in the passive state.

b. Modulation index of vector strength in task-engaged and passive states for fast and slow stimuli separately. (one-sample two-tailed Wilcoxon signed rank with mean 0,  $n=287$ ;  $zval=-4.29$ ,  $p=1.75e-5$  &  $zval=-8.20$ ,  $p=2.36e-16$ ; \*\*\*:  $p<0.001$ ).

c. Modulation index of vector strength in task-engaged and passive states for reference and target stimuli separately. (one-sample two-tailed Wilcoxon signed rank with mean 0,  $n=287$ ;  $zval=-4.95$ ,  $p=7.37e-7$  &  $zval=-7.54$ ,  $p=4.75e-14$ ; \*\*\*:  $p<0.001$ ).

1525



**FigS2.** Reference and target stimuli cannot be discriminated on the basis of population-averaged activity

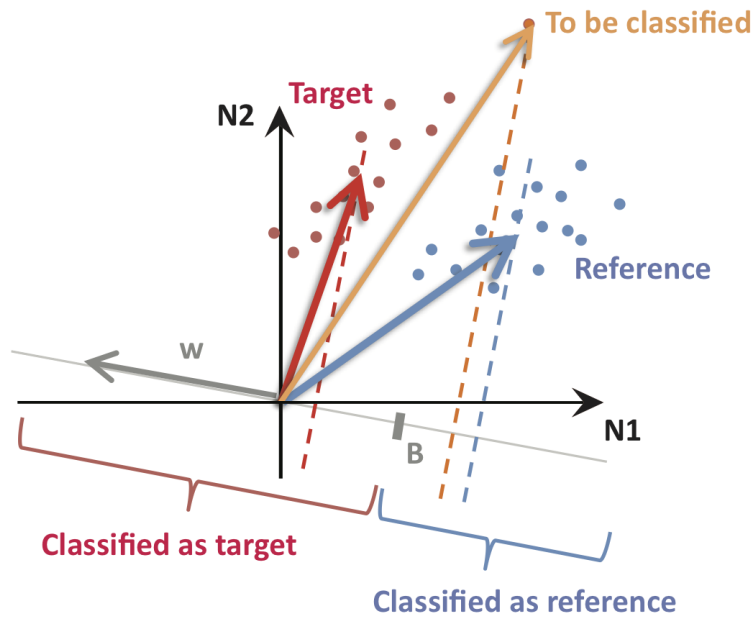
1526  
1527

**Fig S2**

*a. Comparison of average firing rates on log scale in passive (left) and engaged (right) between fast and slow stimuli during the sound. (one-sample two-tailed Wilcoxon signed rank with mean 0,  $n=360$ ;  $zval=-0.53$ ,  $p=0.59$  &  $zval=-0.25$ ,  $p=0.8$ ).*

*b. Accuracy of decoding in engaged and passive state using equal weights for all units. In grey, chance level performance evaluated on label-shuffled trials. Error bars are 1 std over 400 cross-validations*

1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538



**FigS3.** Illustration of binary classifier

1539

**Fig S3**

*Illustration of binary classifier, see materials and methods.*

1540

1541

1542

1543

1544

1545

1546

1547

1548

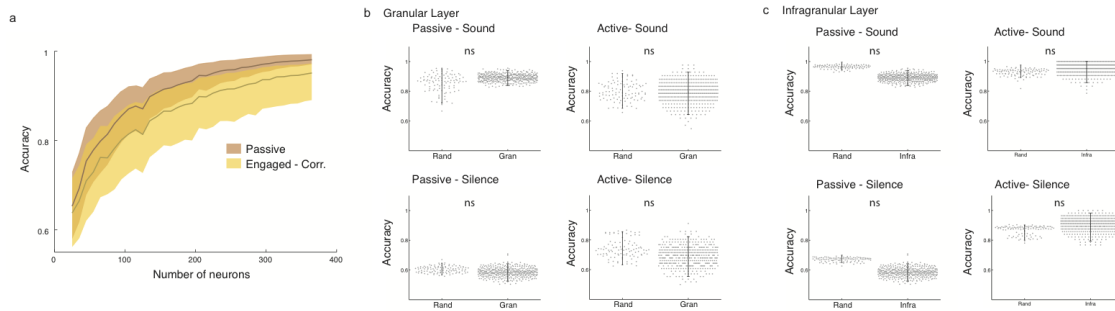
1549

1550

1551

1552

1553



**FigS4.** Properties of the linear classifier

1554

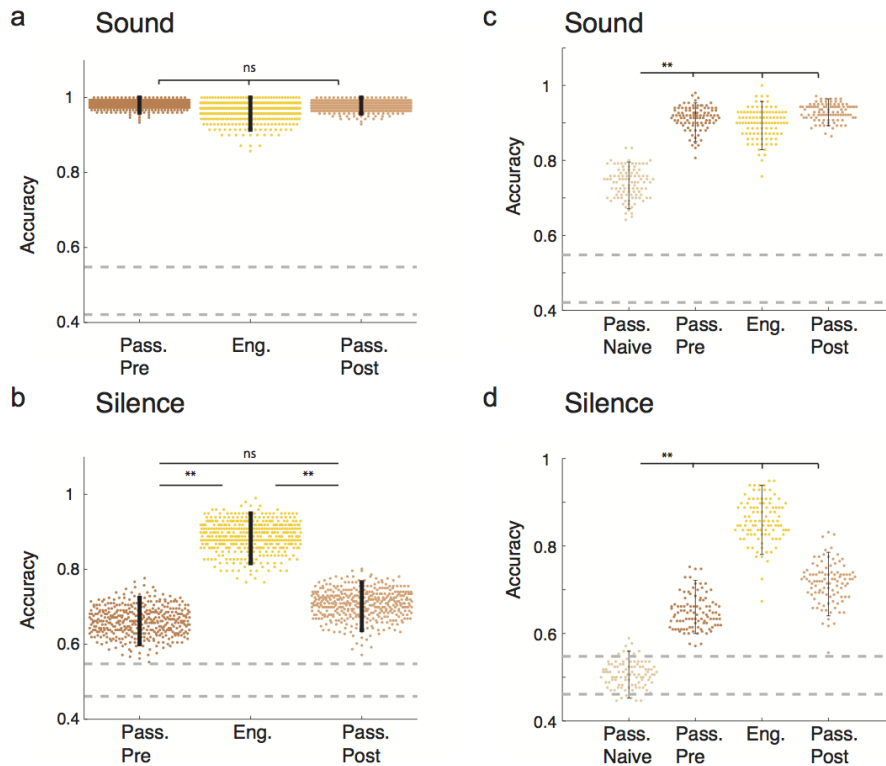
**Fig S4**

*a. Effect of randomly adding units on decoding accuracy during the sound period. Error bar: 95% confidence intervals over 100 random selections of units.*

*b. Units taken from the granular layer only are used for classification and accuracy is compared with the same number (89) of randomly chosen units. Error bars: 95% confidence intervals. (100 sub-sampling procedures, 400 cross validations for accuracy using granular layer units; Bonferonni corrected p-value (8 tests): 0.0063;  $p=0.622$ ,  $p=0.933$ ,  $p=0.624$ ,  $p=0.618$ )*

*c. Same as b but for infragranular layer (273 units). Error bars: 95% confidence intervals. (100 sub-sampling procedures, 400 cross validations for accuracy using granular layer units; Bonferonni corrected p-value (8 tests): 0.0063;  $p=0.0067$ ,  $p=0.51$ ,  $p=0.015$ ,  $p=0.48$ )*

1555



**FigS5.** Comparison of passive sessions before and after behavior

1556

**Fig S5**

a. Comparison of accuracy during the sound period in the passive state before behavior, the task-engaged state and the passive state after behavior. Error bars represent 95% confidence intervals. ( $n=400$  cross validations; pas.pre/eng:  $p=0.45$ , pas.pre/pas.post:  $p=0.74$ , eng/pas.post:  $p=0.58$ ).

b. Comparison of accuracy during the silence period as in a. ( $n=400$  cross validations; Bonferonni corrected  $p$ -value (3 tests): 0.0167; pas.pre/eng:  $p<0.0025$ , pas.pre/pas.post:  $p=0.43$ , eng/pas.post:  $p<0.0025$ ; \*\*:  $p<0.01$ )

c. Comparison of accuracy during the sound period in a naive animal with the passive state before behavior, the task-engaged state and the passive state after behavior in trained animals. For classification, the number of units in the trained animals was downsampled to the same number (222) as those recorded in the naive animal to allow for comparison. Error bars represent 95% confidence intervals. ( $n=100$  cross validations after random downsampling; Bonferonni corrected  $p$ -value (3 tests) : 0.0167; nve/pas.pre, nve/pas.post, nve/eng:  $p<0.0025$ ; \*\*:  $p<0.01$ )

d. Comparison of accuracy during the silence period as in c. ( $n=100$  cross validations after random downsampling; Bonferonni corrected  $p$ -value (3 tests) : 0.0167; nve/pas.pre, nve/pas.post, nve/eng:  $p<0.0025$ ; \*\*:  $p<0.01$ )

1557

1558

1559

1560

1561

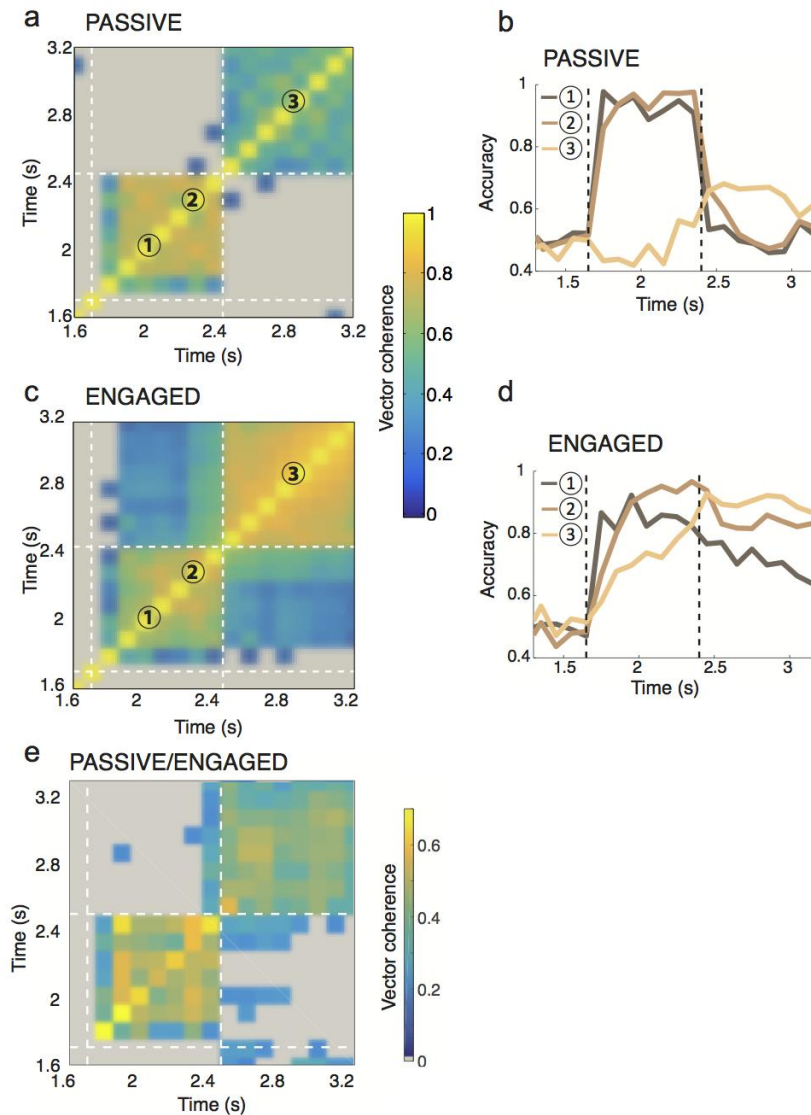
1562

1563

1564

1565

1566



**FigS6.** Comparison of classifiers determined at different time-points and sessions

1567

**Fig S6.**

*a. Classifier evolution in the passive state is shown in colour as the correlation between decoding vectors at one time (y-axis) versus another (x-axis). Squares with below chance correlation values are shown in grey. Here, in the passive state, coding is homogeneous throughout the sound but does not allow for significant decoding in the silent period.*

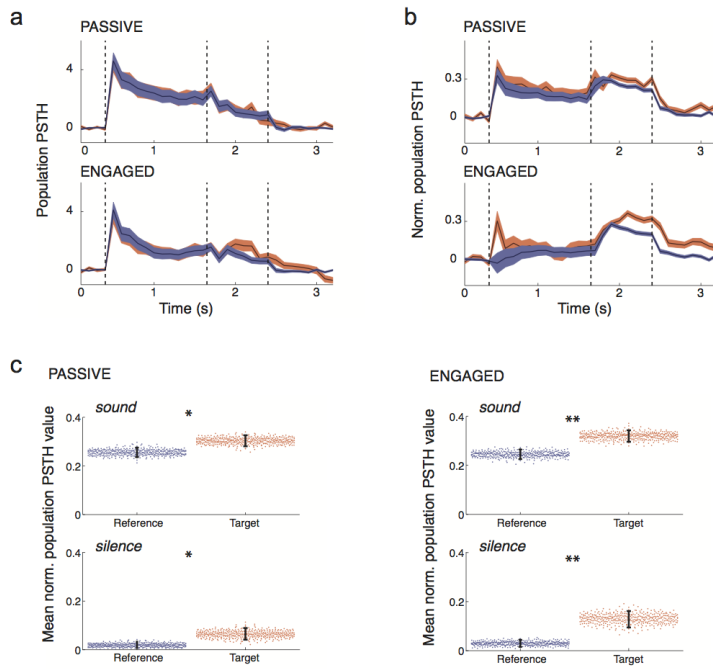
*b. Decoding accuracy in the passive state using a decoder trained on the early (1) or late (2) sound or silence (3) periods. Accuracy is high throughout the sound for both early and late sound training but rapidly falls off during the silence. The decoder trained during the silence is only above chance after the sound has ended.*

*c. Classifier evolution in the task-engaged state as in (a). During the silence, coding is homogeneous.*

*d. As in (b) for the task-engaged state. The decoder trained during the early sound is specific to this period and performs poorly during the silence. Conversely, training late in the sound increases performance during the silence but decreases performance at the beginning of the sound. The accuracy of a decoder trained during the silence ramps up during sound presentation.*

*e. Correlation of passive and engaged decoding vectors throughout the trial. Vectors show stronger similarity during the sound than the silence between states. Note the different color scale, correlation between states is as expected lower than within states.*

1568



**FigS7.** Comparing A1 population-averaged responses to target and reference stimuli

1569

**Fig S7**

*a. Average population PSTH on reference and target trials in the passive and task-engaged states. The PSTH of each neuron is baseline subtracted and then all PSTHs are averaged. Error bars: 95% C.I. after bootstrapping 400 times over all neurons (n=370).*

*b. Average normalized population PSTH on reference and target trials in the passive and task-engaged states. The PSTH of each neuron is baseline subtracted, corrected for the sign of its peak response to reference or target and normalized to its maximal response across states and stimuli. All normalized PSTHs are then averaged. Error bars: 95% C.I. after bootstrapping 400 times over all neurons (n=370).*

*c. Distance of reference and target from baseline after normalization as in (b). Results are shown for both states during the sound or the silence period. Error bars represent 95% confidence intervals. (n=400 cross validations; pass:  $p=0.025$  &  $p=0.025$ , eng:  $p<0.0025$  &  $p<0.0025$ ; \*:  $p<0.05$ ; \*\*:  $p<0.01$ )*

1570

1571

1572

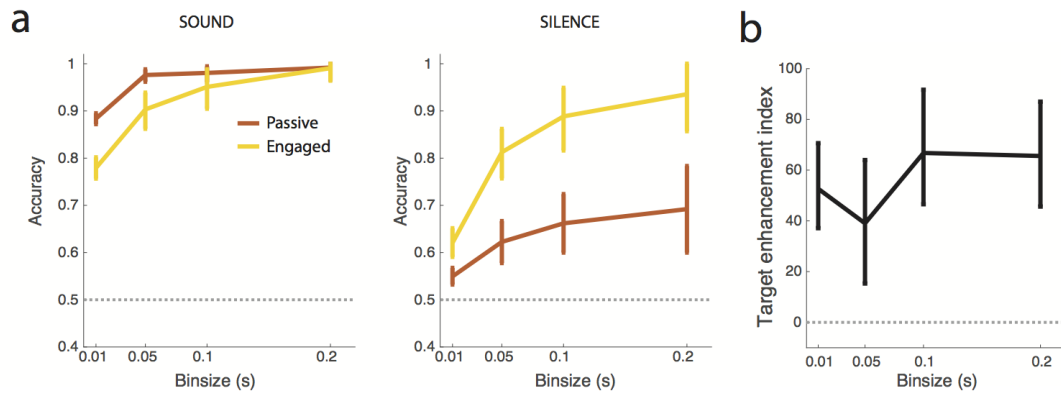
1573

1574

1575

1576

1577



**FigS8.** Robustness of stimulus representation characteristics across a range of time scales

1578

**FigS8.**

*a. Accuracy of decoding during the sound (left) and silence (right) period in passive and engaged states calculated using a classifier determined with time bins of varying size. Error bars represent 95% confidence intervals. (n=400 cross validations)*

*b. Index of target enhancement by task engagement calculated during the sound period using a classifier determined with time bins of varying size. Note that for all time bins the value is significantly greater than 0, indicating a systematic enhancement of target driven encoding in the engaged state. Error bars represent 95% confidence intervals. (n=400 cross validations)*

1579

1580

1581

1582

1583

1584

1585

1586

1587

1588

1589

1590

1591

1592

1593

1594

1595

1596

1597

1598

1599

1600

1601

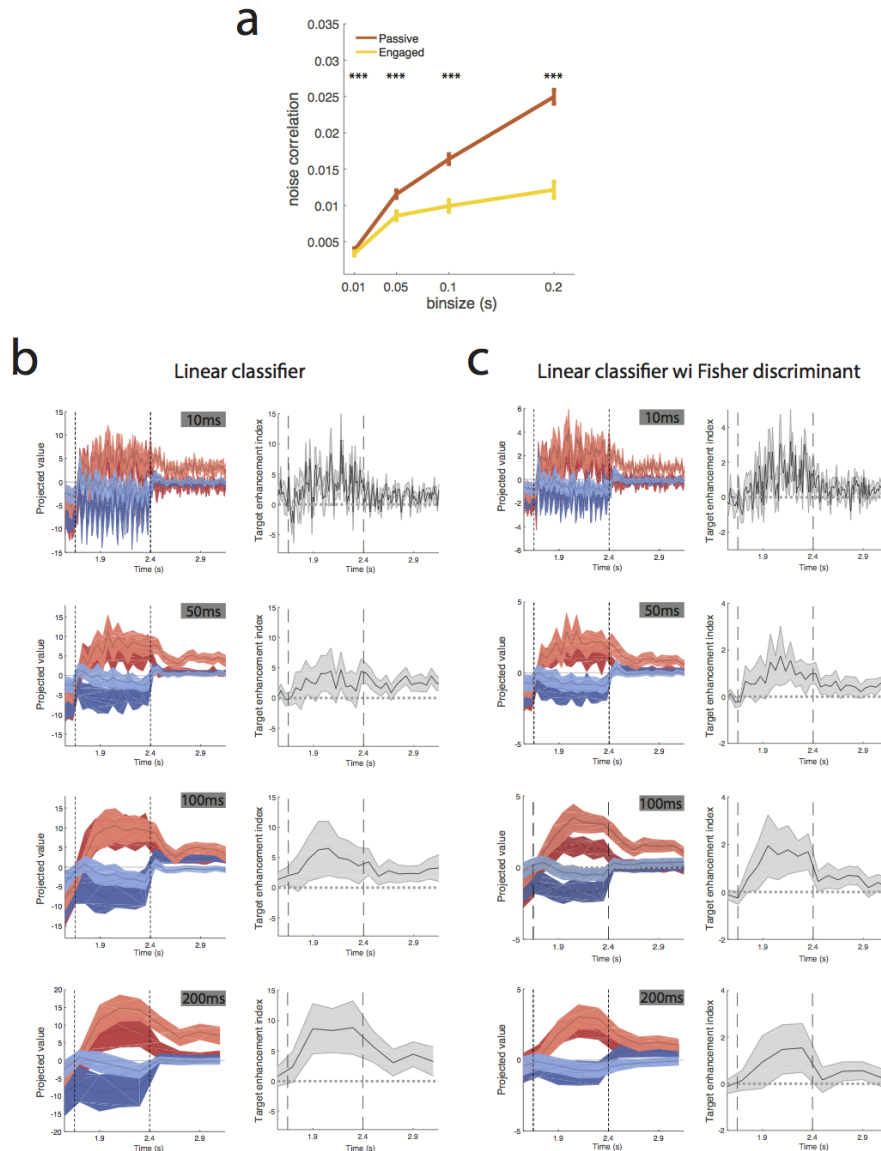
1602

1603

1604

1605

1606



**FigS9.** Reduced noise correlations in the engaged state does not affect enhanced asymmetry at multiple time scales

1607

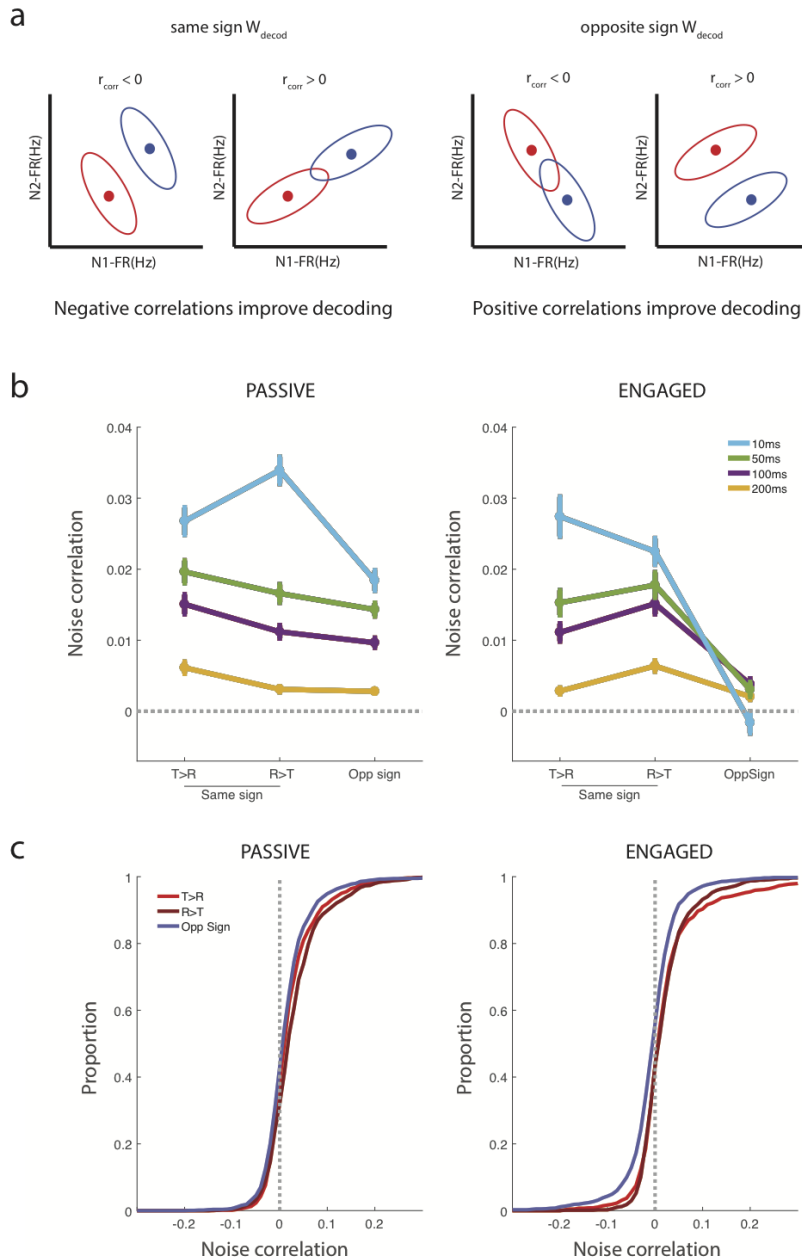
**Fig S9.**

*a. Mean noise correlation in passive and engaged state using time bins of varying duration. . Error bars represent s.e.m over n=3361 pairs(two-sided Wilcoxon signed rank, n=3361 pairs; zval=4.05, p=4.9E-5; zval=7.91, p=2.4E-15; zval=10.33, p=4.9E-25; zval=12.33, p=6.0E-35; \*\*\*:p<0.001)*

*b. Projection onto the decoding axis determined during the sound period of trial-averaged reference (blue) and target (ref) activity during the passive (dark colors) and the active (light colors) sessions and index of target enhancement by task engagement (as in Fig5&8). Time bins of various size were used to define the decoding vector for projection. Note that for easy comparison with the Fisher discriminant analysis, decoding was done on each session individually and then the results for all sessions were averaged.*

*c. As in b, for decoding vector defined using Fisher discriminant analysis.*

1608



**FigS10.** Reduction in noise correlations during task engagement specifically impacts oppositely tuned units

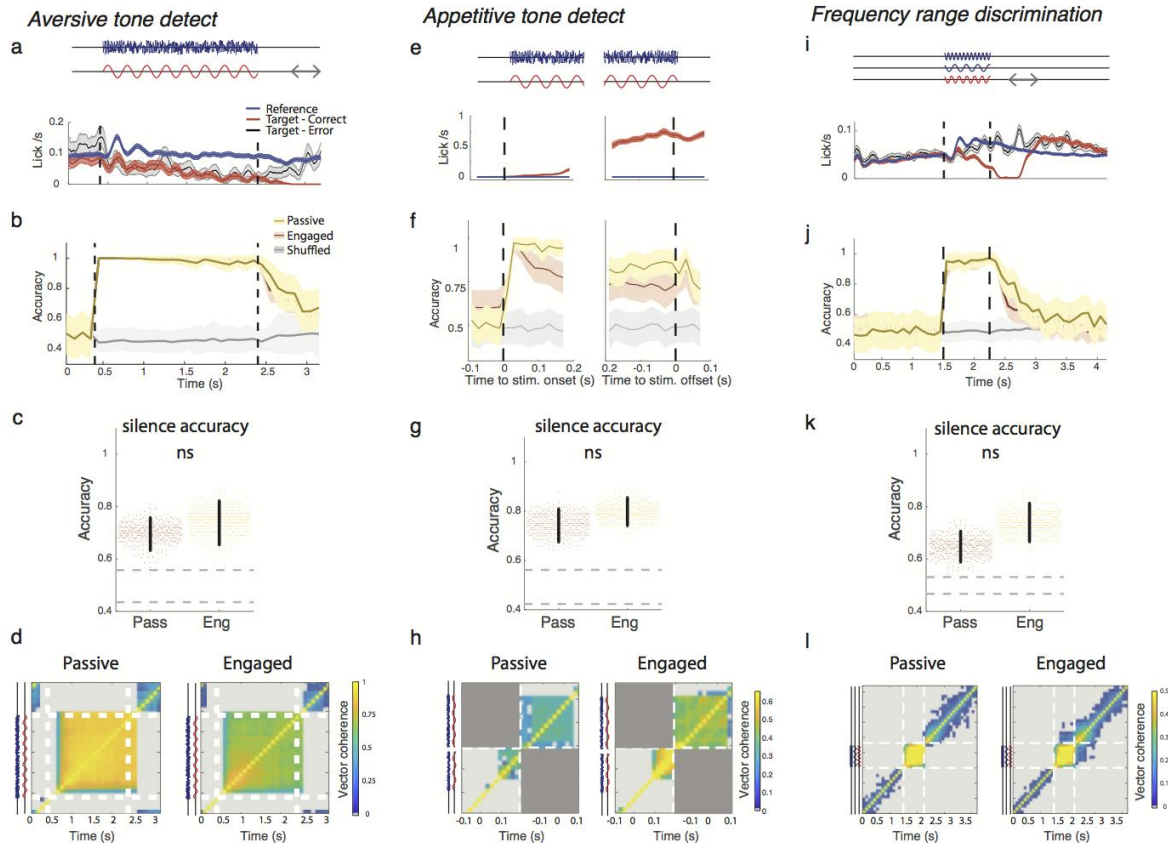
1609  
1610

**Fig S10.**

*a. Schematic illustrating the relationship of ‘signal’ (decoding weight) and ‘noise’ correlations between units. Dots represent the mean target and references responses for two fictive neurons, whereas ellipses show the variance. Negative but not positive noise correlations improve stimulus discrimination for units that have the same sign of decoding weight (ie both are target-preferring or both are reference preferring) whereas the opposite if true of units with opposite sign decoding weights.*

*b. Average noise correlations for units with the same or opposite sign of decoding weight in the passive (left) or engaged (right) state. In the engaged state noise correlations strongly shift towards reduced correlations for all bin sizes used in the analysis.*

*c. Cumulative distribution of noise correlations for units with the same or opposite sign of decoding weight in the passive (left) or engaged (right) state. Note that the distributions are similar in the passive state whereas in the active state there is a clear shift of the noise correlations for units of opposite decoding weight sign towards lower values. There is a clear enhancement of negative correlation values.*



**FigS11.** Task structure and decoding of reference/target activity in a range of auditory go/no-go tasks

1611

**Fig S11.**

Three different tasks are considered: aversive tone detect (a-d), appetitive tone detect (e-h) and frequency range discrimination (i-l). Note that all analysis in this figure is done after excluding lick-responsive units for these tasks using the method described in Fig 4.

a, e, i. Top: Schematic of trial structure illustrating reference and target trials. Gray arrows show response window for the aversive tasks. Bottom: Licking frequency during correct target (red), reference (blue) and target error (gray) trials. Error bars are s.e.m over all trials.

b, f, j. Accuracy of stimulus classification in passive and engaged states. In grey, chance level performance evaluated on label-shuffled trials. Error bars represent 1 std calculated over 400 cross-validations.

c,g,k. Mean classifier accuracy during the post-sound silence period in passive and engaged conditions. Gray dotted lines give 95% confidence interval of shuffled trials. Error bars represent 95% confidence intervals. Note that accuracy is systematically above chance level in both conditions but does not change between the passive to the engaged state. ( $n=400$  cross validations;  $p=0.21,0.18,0.055$ )

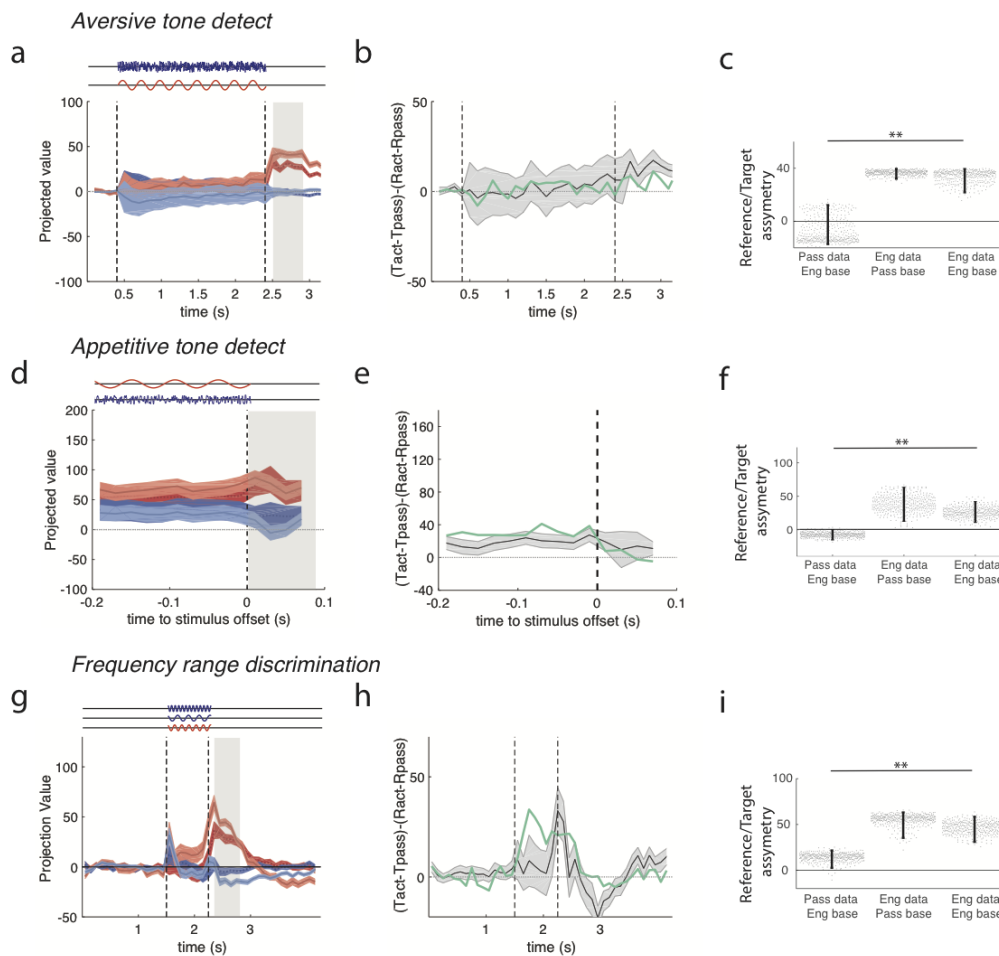
d,h,l. Classifier evolution in the passive (left) and engaged (right) state is shown in color as the correlation between decoding vectors at one time (y-axis) versus another (x-axis). Squares with below chance correlation values are shown in grey. For the appetitive tone detect task the overlap between sound onset and sound offset periods is not calculated as the difference in trial durations causes different overlaps in time on a trial to trial basis between the two. Note that the sound and silence periods in all tasks rely on different decoding vectors and in the case of the frequency range discrimination task, there is a progressive shift in the engaged state between decoders.

1612

1613

1614

1615



**FigS12.** Asymmetric encoding of target and reference stimuli in a range of auditory go/no-go tasks during the post-sound silence

1616

**Figure S12**

*a,d,g* Projection of onto the decoding axis determined during the post-sound silence period of trial-averaged reference (blue) and target (red) activity during the passive (dark colors) and the active (light colors) sessions. A baseline value computed from pre-stimulus spontaneous activity was subtracted for each neuron, so that the origin corresponds to the projection of spontaneous activity (shown by black line). Note that there is a tendency for the target-driven activity to be further from the baseline in the active state and/or the reference-driven activity to be closer. The periods used to construct the decoding axis are shaded in gray. Error bars represent 1 std calculated using decoding vectors from cross-validation ( $n=400$ ).

*b,e,h* Index of target enhancement by task engagement based on projections using the decoding axis determined during post-sound silence. In green same index instead giving the same weight to all units. The difference between the green and black curved indicates that the change in asymmetry induced by task engagement cannot be detected using the population averaged firing rate alone. Error bars represent 1 std calculated using decoding vectors from cross-validation ( $n=400$ ).

*c,f,i* Comparison of reference/target asymmetry for evoked responses in different states during the post-sound silence compared to different baselines given by passive or engaged spontaneous activity. Reference/target asymmetry is the difference of the distance of target and reference projected data to a given baseline. We examine three cases: (i) passive evoked responses, distances calculated relative to engaged spontaneous activity; (ii) engaged evoked responses, distances calculated relative to passive spontaneous activity; (iii) engaged evoked responses, distances calculated relative to engaged spontaneous activity. In all three cases, the engaged decoding axis was used for projections. Error bars represent 95% confidence intervals. ( $n=400$  cross validations; Aversive Tone detect:  $p(\text{col1,col3}) < 0.0025$  &  $p(\text{col2,col3}) = 0.92$ ; Appetitive tone detect;  $p(\text{col1,col3}) < 0.025$  &  $p(\text{col2,col3}) = 0.94$ ; Frequency range discrimination:  $p(\text{col1,col3}) < 0.0025$  &  $p(\text{col2,col3}) = 0.9$ ; \*\*:  $p < 0.01$ ).