

Title: Population coding of time-varying sounds in the non-lemniscal Inferior Colliculus

Abbreviated title: Population coding in shell IC

Kaiwen Shi^{1*}, Gunnar L. Quass^{1*}, Meike M. Rogalla¹, Alexander N. Ford¹, Jordyn E. Czarny¹,
Pierre F. Apostolides^{1,2}

1. Kresge Hearing Research Institute, Department of Otolaryngology — Head & Neck Surgery

2. Department of Molecular & Integrative Physiology

*Equal Contribution

University of Michigan Medical School, Ann Arbor, MI, 48109

Correspondence: Pierre F. Apostolides, Ph.D. piaposto@med.umich.edu

Number of pages: 63

Number of figures: 9

Number of tables: 1

Word counts:

Abstract: 247

Significance Statement: 115

Introduction: 719

Discussion: 979

Acknowledgments: Funding was provided by the Whitehall Foundation, Hearing Health Foundation, and NIH/NIDCD R01DC019090 to PFA and Magnificent Michigan Undergraduate Research Fellowship to KS. The authors thank Dr. Anahita Mehta for helpful comments.

Author Contributions: KS, GLQ, MMR, and PFA designed the research. KS, GLQ, ANF, JEC, and PFA conducted the research. KS, GLQ, and MMR analyzed the data. KS, GLQ and PFA wrote the paper.

Conflict of interest statement: The authors report no competing interests.

1 Abstract

2 The inferior colliculus (IC) of the midbrain is important for complex sound processing, such as
3 discriminating conspecific vocalizations and human speech. The IC's non-lemniscal, dorsal
4 "shell" region is likely important for this process, as neurons in these layers project to higher-
5 order thalamic nuclei that subsequently funnel acoustic signals to the amygdala and non-primary
6 auditory cortices; forebrain circuits important for vocalization coding in a variety of mammals,
7 including humans. However, the extent to which shell IC neurons transmit acoustic features
8 necessary to discern vocalizations is less clear, owing to the technical difficulty of recording
9 from neurons in the IC's superficial layers via traditional approaches. Here we use 2-photon Ca^{2+}
10 imaging in mice of either sex to test how shell IC neuron populations encode the rate and depth
11 of amplitude modulation, important sound cues for speech perception. Most shell IC neurons
12 were broadly tuned, with a low neurometric discrimination of amplitude modulation rate; only a
13 subset were highly selective to specific modulation rates. Nevertheless, neural network classifier
14 trained on fluorescence data from shell IC neuron populations accurately classified amplitude
15 modulation rate, and decoding accuracy was only marginally reduced when highly tuned neurons
16 were omitted from training data. Rather, classifier accuracy increased monotonically with the
17 modulation depth of the training data, such that classifiers trained on full-depth modulated
18 sounds had median decoding errors of ~0.2 octaves. Thus, shell IC neurons may transmit time-
19 varying signals via a population code, with perhaps limited reliance on the discriminative
20 capacity of any individual neuron.

21 Significance Statement

22 The IC's shell layers originate a “non-lemniscal” pathway whose first- and second-order targets
23 are thought important for perceiving conspecific vocalizations and human speech. However,
24 prior studies suggest that individual shell IC neurons are broadly tuned and have high response
25 thresholds, implying a limited reliability of efferent signals. Here we use Ca^{2+} imaging to test
26 how shell IC neurons encode amplitude modulation, a key sound cue for speech perception and
27 stream segregation. We show that the rate and depth of amplitude modulation is accurately
28 represented in the ensemble activity of shell IC neuron populations. Thus, downstream targets
29 can read out a sound's temporal envelope from a distributed rate code transmitted by populations
30 of broadly tuned neurons.

31 **Introduction**

32 The “amplitude modulation” (AM) of a sound’s envelope is a hallmark of vocal communication
33 across species, and supports important cognitive processes such as stream segregation and
34 intelligibility of syllabic features (Drullman et al., 1994; Shannon et al., 1995; Zeng et al., 2005;
35 Gallun and Souza, 2008; Elliott and Theunissen, 2009; Koumura et al., 2019, 2023). Despite this
36 importance, how the brain reliably encodes AM sounds, and thus may provide a key building
37 block of receptive language, remains poorly understood.

38 The inferior colliculus (IC) is an evolutionarily conserved midbrain structure that plays a pivotal
39 role in AM perception (Champoux et al., 2007), and the first major central auditory structure
40 where neuronal responses to AM sounds change from a temporal- to a rate-based representation
41 (Hewitt and Meddis, 1994; Lorenzi et al., 1995; Krishna and Semple, 2000; Nelson and Carney,
42 2007; Tan and Borst, 2007; Ter-Mikaelian et al., 2007; Dicke et al., 2007; Geis and Borst, 2009;
43 Wang et al., 2022). Indeed, although neurons in sub-tectal brainstem nuclei phase-lock their
44 firing rates to AM fluctuations, mean spike rates are often insensitive to the temporal
45 characteristics of the sound envelope (Frisina et al., 1990; Rhode and Greenberg, 1994; Zhao and
46 Liang, 1994). By contrast, mean spike rates of IC neurons consistently and dramatically vary
47 depending on a sound’s AM pattern, with specific IC neurons often displaying sharp AM rate
48 selectivity (Langner and Schreiner, 1988; Krishna and Semple, 2000; Nelson and Carney, 2007).
49 Compared with a temporal code which requires sampling over multiple period cycles, such rate
50 coding across spatially distributed neurons might facilitate more efficient representation of sound
51 features (Bendor and Wang, 2007; Wang et al., 2008) and drive behavior (Bagur et al., 2023).

52 The IC is also the first bifurcation site of primary and “higher-order” acoustic pathways. Neurons
53 in the IC’s central nucleus project to the primary auditory thalamus, the ventral medial geniculate
54 body (MGB). By contrast, neurons located in the surrounding dorsomedial and lateral “shell” IC
55 nuclei project to the higher-order dorsal and medial MGB (Oliver and Morest, 1984; Winer et al.,
56 1998; Mellott et al., 2014; Xiong et al., 2015; Chen et al., 2018; Cai et al., 2019; Oberle et al.,
57 2022). This divergence is notable, as higher-order MGB nuclei relay acoustic signals to the
58 amygdala, non-primary auditory cortices, and the striatum (LeDoux et al., 1991; Bordi and
59 LeDoux, 1994; Lee and Winer, 2008; Ponvert and Jaramillo, 2019). These forebrain circuits are
60 important for processing conspecific vocalizations and human speech (Gadziola et al., 2012;
61 Grimsley et al., 2013; Gadziola et al., 2016; Hamilton et al., 2021), as well as orchestrating
62 behavioral responses to ethologically relevant sounds (Miura et al., 2020; Carcea et al., 2021;
63 Valtcheva et al., 2023).

64 Despite prominent connections with behaviorally relevant circuits, little is known about how
65 shell IC neurons respond to AM sounds; most studies of AM coding have been conducted in the
66 central IC. This knowledge gap may be due to shell IC neurons’ high response thresholds, loose
67 tonotopy, and variable response fidelity compared to central IC neurons (Syka et al., 2000;
68 Lumani and Zhang, 2010; Barnstedt et al., 2015; Parras et al., 2017; Wong and Borst, 2019;
69 Chen and Song, 2019). Additionally, shell IC neurons’ precarious location at the superficial
70 tectal layers complicates the use of single neuron recording techniques to study a variably active
71 neuronal population. Rather, a tacit assumption is that non-lemniscal auditory system eschews
72 high-fidelity representations of acoustic features in favor of coarser signals potentially related to
73 a sound’s contextual novelty (Ayala et al., 2015; Parras et al., 2017) or behavioral salience (Olds
74 et al., 1972; Edeline and Weinberger, 1992; Poremba and Gabriel, 2001; Leppla et al., 2023).

75 Here we use 2-photon Ca²⁺ imaging to study how mouse shell IC neurons respond to
76 sinusoidally amplitude modulated (sAM) sounds. We find that all the major sAM tuning
77 properties previously described in the central IC – low-pass, high-pass, band-pass, and band-
78 reject – are also found in the shell IC. Increasing the sAM depth monotonically enhanced
79 activity, indicating limited mixed selectivity to specific combinations of sAM rate and -depth.
80 Although individual shell IC neurons often displayed low neurometric selectivity for sAM
81 sounds, decoder algorithms trained on shell IC fluorescence data could accurately classify sAM
82 features with high reliability. Thus, a population code enables the IC’s non-lemniscal efferents to
83 can transmit detailed information regarding time-varying acoustic features.

84 Materials & Methods

85 Surgeries:

86 Experiments were approved by Michigan Medicine’s IACUC and performed in accordance with
87 NIH’s guide for the care and use of laboratory animals. Surgical procedures were performed
88 under aseptic conditions with 5-8-week-old male and female C57/Bl6J mice ordered from
89 Jackson Labs (N = 9, 4 female and 5 male) or offspring of VGAT-ires-cre × Ai14 breeders in our
90 colony (N = 4, 2 female and 2 male). Mice were anesthetized with 4-5% isoflurane delivered in
91 O² and mounted in a stereotaxic apparatus (David Kopf Instruments) and injected with carprofen
92 as pre-surgical analgesic. Body temperature was maintained near 37 C° with a feedback-
93 controlled homeothermic blanket (Harvard Apparatus). Isoflurane concentration was
94 subsequently reduced to 1.5-3 %, the scalp was cleaned with iodine solution (betadine), shaved,
95 and subsequently removed. The periosteum was carefully removed and the exposed skull was
96 etched with a sharp scalpel. A ~2 mm circular craniotomy was carefully drilled over the left IC
97 using a Foredom microdrill (0.9 mm caudal and 1.0 mm lateral from the lambdoid sutures),

98 employing multiple rounds of short (<10 s) drilling, interleaved with ~1-2 min application
99 periods of chilled phosphate buffered saline (PBS) to the skull. Care was taken so as to not
100 damage the transverse sinus bordering the IC and superior colliculus during the surgery.
101 Following removal of the skull, 12.5-25 nL of virus solution was injected at 3-6 sites in the IC
102 (~100 nL total volume). A custom-made cranial window comprised of three stacked 2 mm
103 diameter coverslips adhered to a 4 mm diameter coverslip (Potomac Glass) was inserted into the
104 craniotomy and held in place with gentle pressure. The cranial window was bonded to the skull
105 with cyanoacrylate glue (Loctite) and dental cement (Ortho-Jet). The skull was then covered in a
106 thin layer of glue and a custom titanium headbar was cemented onto the skull. Mice received a
107 post-operative injection of buprenorphine (0.03 mg/kg s.c., Par Pharmaceuticals) for analgesia
108 and 24 hours later, a second dose of carprofen (5 mg/kg, s.c., Spring Meds). Mice were
109 monitored for 7-10 days following surgery for any sign of discomfort.

110 **2-photon Ca²⁺ imaging:**

111 Beginning 10-14 days following surgery, mice were handled by the experimenter and acclimated
112 to increasingly prolonged durations of head-fixation while sitting quietly in a custom acrylic
113 glass tube. For data collection, mice were head-fixed under a Janelia MIMMs style microscope
114 (Sutter Instruments) and 2-photon excitation was delivered at 920 nm through a Nikon 16x
115 objective focused axially through the cranial window onto the superficial layers of the left IC
116 using a tunable laser (Coherent Chameleon). Peak laser power measured at the objective was
117 typically 20-40 mW and did not exceed 60 mW. GCaMP and tdTomato fluorescence were
118 collected using GaASP photomultiplier tubes (Hamamatsu). Frame scans (512 x 512 pixels, 30
119 Hz frame rate) were collected for 5-6 s on each trial, with a 3-5 s inter-trial interval where no
120 laser light was delivered to the brain.

121 **Stimulus presentation:**

122 Acoustic stimuli were generated using a Tucker-Davis Technologies RZ6 and delivered in free
123 field at 65-70 dB SPL via a calibrated electrostatic speaker (TDT ES1) positioned ~10 cm from
124 the mouse's right ear. Stimuli with different sAM rates and depths were delivered in pseudo-
125 random order once every 8-11 s. We typically delivered 10 repetitions of each stimulus when
126 imaging a single FOV, which required 40-80 min to complete. 1-2 FOVs were typically imaged
127 from a single mouse in each session, which typically required 1-2 hours of total head-fixation
128 time per session. Subjects underwent multiple imaging sessions over the course of 2-3 weeks.

129 **Data analysis:**

130 Raw movies were motion-corrected using non-rigid registration, and regions of interests (ROIs)
131 corresponding to individual cell bodies were generated to extract the raw fluorescence signal
132 from somata and surrounding neuropil using Suite2p (Pachitariu et al., 2016). Fluorescence
133 changes were measured by subtracting the corresponding neuropil signal (scaled between 70-
134 100 % based on recording quality) on each trial and calculating the change in fluorescence as ($F -$
135 F_0)/ F_0 ($\Delta F/F$). In these experiments, F_0 represents the mean fluorescence intensity during the 0.5
136 - 1 s baseline period prior to sound onset on each trial.

137 Analyses presented in this paper are performed only on ROIs displaying a statistically significant
138 sound response, as determined using the 'signal autocorrelation' bootstrapping procedure of Geis
139 et al., (2011) and Wong and Borst (2019) and for each ROI in an FOV, the raw $\Delta F/F$ traces
140 during, and 1 s following, sound presentation were correlated between individual repetitions of
141 the same stimulus presentation. We calculated a mean Pearson's ρ value across all possible pairs
142 of trials of the same stimulus presentation, thus providing a metric of how reliably the particular
143 sound stimulus drives a response in the neuron. We then generated "shuffled" datasets of

144 randomly chosen sequential frames from the fluorescence time series across all trials of the same
145 stimulus presentation. The length of these segments, and the number of pairs of trials, were
146 always equal to the length of fluorescence traces and number of stimulus presentations used to
147 calculate the true signal autocorrelation described above. In addition, the specific fluorescence
148 time series was included in only a single randomly chosen segment. We calculated the signal
149 autocorrelation for these shuffled data sets were and the process was bootstrapped 10,000 times.
150 A p-value was then calculated for the ROI, defined as the percentage of shuffled bootstrapped
151 iterations whose mean rho value exceeded that of the true data after applying the Bonferroni-
152 Holm correction.

153 For measures of peak $\Delta F/F$, we averaged all trials corresponding to the same sound stimulus and
154 calculated the mean signal centered ± 1 frames around the local absolute maximum of the
155 fluorescence change during sound presentation.

156 **Lifetime sparseness:**

157 Lifetime sparseness was used to calculate the selectivity of single shell IC neurons to sAM rates
158 for fully modulated and unmodulated sounds,

$$159 S = \frac{1}{1 - 1/N} \left(1 - \frac{\left(\sum_{j=1}^N r_j / N \right)^2}{\sum_{j=1}^N r_j^2 / N} \right),$$

160 where N is the number of different sAM rates and r_j is the peak of the neuronal response to sAM
161 rate j during sound presentation. Lifetime sparseness was calculated for every individual neuron.

162 **Neurometric sensitivity index (d-prime)**

163 To assess the shell IC neurons' ability to separate different sound stimuli, we computed d-prime
164 for all sound-responsive neurons. The neurometric d-prime for a single neuron is defined as

165 $d\text{-prime} = z(Hit\ rate_d) - z(False\ alarm\ rate_d)$,

166 where z is the inverse of the normal cumulative function, and d is the sAM depth of the sound.

167 We define a trial to be a hit if a neuron's response at its preferred sAM rate exceeded 3 times the

168 standard deviation during its baseline (1s before sound onset). Similarly, a trial is defined as a

169 false alarm if that neuron's response at non-preferred sAM rates exceeded 3 times the standard

170 deviation during the baseline. In addition, the standard deviation σ of the difference of

171 neurometric d -primes between 100% and 20% sAM depth for the stimuli at the preferred rate, of

172 all neurons was calculated. If this difference was larger or smaller than 2σ or -2σ , respectively,

173 its d -prime was defined to have either an increasing or decreasing trend.

174 **Convolutional neural network (CNN) decoder:**

175 A 3D-matrix containing neural population Ca^{2+} fluorescence activity from 38 sessions from 13

176 mice (Neurons \times Frames \times Trials) was passed to a CNN decoder (MATLAB_R2021b). The

177 decoder was composed of an input layer, a convolutional layer, a max-pooling layer, a fully

178 connected layer, and an output layer (classification or regression layer). A dropout layer and L2

179 regularization were added to prevent overfitting. Below we provide the details of operations of

180 these building blocks.

181 *Convolutional Layer:*

182 The convolutional layer, consisting of k kernels, performs a linear operation to extract features

183 from the input. The following parameters mainly define the convolutional layer:

184 s_k, s_p, s_s, k , and \mathbf{W} , where s_k, s_p, s_s are the size of the kernels, the padding, and the stride, k is

185 the number of kernels, and \mathbf{W} is the weight array of the kernels. The input \mathbf{X} for the

186 convolutional layer is a temporal representation of each neuron's Ca^{2+} fluorescence, expressed as

187 a 2D array (h_{in}, w_{in}), where the height h_{in} is the number of neurons in each imaging session, the
188 width w_{in} is the evaluation time (from sound onset to 1s or 0.5s after sound offset, depending on
189 sound duration), and each entry contains the normalized fluorescence value for a given neuron at
190 a given time. K filtered feature maps are generated after each kernel is convolved with the input.

191 The output of the convolutional layer $\mathbf{Y} (\frac{h_{in}-s_k+s_p}{s_s} + 1, \frac{w_{in}-s_k+s_p}{s_s} + 1, k)$ is expressed as:

192
$$\mathbf{Y}[I, j, k] = \sum_{h,w} \mathbf{W}[h, w, k] \odot \mathbf{X}[i + h, j + w],$$

193 where \odot denotes element-wise multiplication. In our analysis, the size of the kernels is set to be
194 6×6 and the number of kernels is 32. A non-linearity is then added to the output of the
195 convolutional layer through the rectified linear unit (ReLU) activation function, which is defined
196 as:

197
$$ReLU(y) = \begin{cases} y & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases}$$

198 This nonlinearity enables a neural network to learn and operate with nonlinear functions. The
199 size of the stride was 1×1 and we used zero-padding to avoid input/output size-discrepancies.

200 *Max-Pooling Layer*

201 The pooling layer downsamples its input via operation in a $q \times q$ region. Taking the input of size
202 (h_{in}, w_{in}) , a max-pooling layer returns a 2D output array $(\frac{h_{in}}{q}, \frac{w_{in}}{q})$ by taking the maximum value
203 within the $q \times q$ region. Here, the size of max-pooling is 3×3 with a 1×1 stride.

204 *Fully Connected (FC) Layer*

205 The FC layer first flattens its input $\mathbf{X} (h_{in}, w_{in})$ into a one-dimensional array (n_{out}). This is
206 achieved by reducing the input \mathbf{X} to an array $\mathbf{X}_{flat} (h_{in} \times w_{in})$, and multiplying X_{flat} with its

207 weight matrix \mathbf{W} . L_2 regularization is incorporated in the FC layer, and the FC layer is followed
208 by the output layer (either classification or regression).

209 *Classification Layer*

210 For classification tasks, the SoftMax layer transfers the outputs of the FC layer into a probability
211 distribution of N labels by a normalized exponential function. Following the SoftMax layer, the
212 classification layer adopts the cross-entropy loss function,

$$213 \quad loss = -\frac{1}{N} \sum_i^N \sum_j^K y_{i,j} \ln \hat{y}_{i,j},$$

214 where N is the size of the training set, K is the number of classes, \hat{y} is the probability generated
215 by the SoftMax layer and y indicates the ground truth.

216 *Regression Layer*

217 For regression tasks, the FC layer is followed by a regression layer, which adopts the mean-
218 square-error (MSE) as a loss function,

219

$$220 \quad MSE = \sum^N \frac{(\hat{y}_i - y_i)^2}{N},$$

221 where \hat{y}_i is the prediction generated by the network, y_i is the ground truth, and N is the number
222 of trials.

223 *Dropout Layer*

224 The dropout layer randomly sets 50 % of the weights of the FC layer to zero for one training
225 batch, while the rest remains untouched. This operation prevents overfitting of a neural network

226 by avoiding complicated co-adaptations of feature detectors in the training data (Hinton et al.,
227 2012).

228 *L₂ Regularization*

229 Regularization is one of the most effective ways to prevent overfitting. An *L₂* regularizer
230 penalizes the loss functions by introducing a term $\lambda \sum_{i=0}^N w_i^2$ to the loss function, where w
231 represents the weight and λ represents a coefficient determining the scale of regularization.
232 (Cortes et al., 2012)

233 **Network training and evaluation**

234 The neural network was trained to decode sAM sound parameters using neural population
235 activity by stochastic gradient descent. We used the ADAM optimization algorithm (Kingma and
236 Ba, 2014) with a learning rate of 0.0001, L₂ regularization factor 1.2-1.5, mini-batch size of 64,
237 and epoch 259-519. The classifier was trained using 80% of the dataset, and the remaining 20%
238 was held out as a validation set, with training and testing sets containing equal proportions of
239 different stimuli. The number of iterations for the network is $\frac{Trials*Epoch}{Batch\ Size}$, and for classification
240 tasks, the training accuracy was validated every 20 iterations. The resulting testing accuracy was
241 averaged over all animals and imaging sessions. For the regression task (high-resolution sAM
242 rate decoding), the decoding error is expressed in octaves,

243
$$\text{Decoding error} = \log_2 \left(\frac{\text{Prediction}}{\text{Ground Truth}} \right).$$

244 We trained the CNN decoder to perform different decoding tasks. For the first task with multiple
245 pairs of sAM depths and -rates, we trained the decoder to classify the joint combination of sAM
246 depths and -rates (Figure 4B), sAM rates under a given depth (Figure 4C, D), and sAM depths
247 under a given rate (Figure 6). The classifier was further trained on an exclusive sAM depth or -

248 rate to decode sAM rates or -depths, respectively. This trained classifier was then tested on data
249 with other sAM depths or -rates to predict rates or -depths, and the resulting decoding accuracy
250 was fitted by a gaussian model (Figure 7B). Validation of our model was performed by assessing
251 the confusion matrix, feeding the classifier a shuffled neuron order in the input and receiver
252 operating characteristic (ROC, Figure 4E). For the second experiment with more narrowly
253 spaced sAM rates and 100% sAM depth, we trained the CNN decoder to regress, not classify,
254 sAM rates (Figure 9D-E).

255 **Classifier network d-prime**

256 To evaluate the sAM rate classification performance, we calculated the d-prime for the decoding
257 performance of the classifier network. We define the decoder d-prime the same as the
258 neurometric d-prime, calculating the hit rate and false alarm rate based on the confusion matrix
259 from every imaging session. The d-prime was calculated via the one-vs-rest manner, which
260 assumes that the positive class only includes a specific sAM rate, while the negative class
261 contains the rest of sAM rates.

262 **Pattern correlation**

263 To measure similarity of neural population activities associated with different sAM sounds, we
264 adapted the pattern correlation described in the work of Friedrich and Laurent (2001) and
265 population vector correlation in the work of Deitch et al. (2021) as follows. We first trial-
266 averaged responses of single neuron associated different combinations of sAM depths and -rates.
267 Pattern correlation of trial-averaged neural population activity in shell IC between two different
268 sAM depths or -rates calculated on a per-frame basis using Pearson's correlation coefficient.

269 **t-distribution stochastic neighbor embedding (t-SNE) clustering.**

270 The t-SNE analysis, a dimensionality reduction technique where the neural population activity is
271 represented as a single data point in a two-dimensional map, was performed according to the
272 framework of van der Maaten and Hinton (2008). The local proximity of data points in the map
273 captures the similarity of data in the original dimension and t-SNE can thus cluster the data
274 points.

275 In our case, the data contained N vectors: x_1, \dots, x_n , where each x is a neural population activity
276 vector containing the mean $\Delta F/F$ of every neuron during sound presentation. t-SNE transforms
277 the Euclidean distance of a pair of vectors x_i and x_j in the high dimensional data into a
278 conditional probability $p_{i|j}$, representing the similarity between two vectors,

$$279 \quad p_{i|j} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}$$

280 where the variance σ_i^2 is determined such that, the beforehand determined perplexity \mathcal{P} satisfies

$$281 \quad \mathcal{P} = 2^{-\sum_{i \neq j} p_{i|j} \log_2 p_{i|j}}.$$

282 For our study, the perplexity ranges from 10 to 50. The t-SNE algorithm then constructs a
283 mapping of a low dimensional representation y_i and y_j , which is determined via a random walk
284 approach, to the high dimensional data x_i and x_j . The similarity between the low dimensional
285 counterparts is represented by the conditional probability $q_{i|j}$ of y_i and y_j is given by the student
286 t-distribution with one degree of freedom,

$$287 \quad q_{i|j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}},$$

288 The aim of t-SNE is to minimize the mismatch between $q_{i|j}$ and $p_{i|j}$, so that the low dimensional
289 counterpart of \mathbf{y}_i and \mathbf{y}_j can optimally model the similarity between the high dimensional data
290 \mathbf{x}_i and \mathbf{x}_j . Thus, the t-SNE adopts the Kullback-Leibler divergence of $q_{i|j}$ and $p_{i|j}$ as a loss
291 function C,

$$292 \quad C = \sum_i \sum_j p_{i|j} \log \frac{p_{i|j}}{q_{i|j}},$$

293 and adjusts the low dimensional representation by minimizing C via adaptive gradient descent.
294 The optimized low dimensional counterpart $\mathbf{y}_1, \dots, \mathbf{y}_n$, two dimensional in our case, was
295 visualized. The values in the axes of two dimensions given by t-SNE do not have a specific
296 meaning, but only capture the similarity in the original data.

297 **Statistics:**

298 Statistical assessments were performed using MATLAB_R2021b and GraphPad Prism 9. For all
299 within- and between-group comparisons, data were tested for normality using Lilliefors' test.
300 Parametric approaches were adopted for data following a normal distribution, while non-
301 parametric approaches were adopted for data not normally distributed. The alpha level is
302 corrected for multiple comparisons using the Bonferroni-Holm method. For repeated-measures
303 ANOVA, Geisser-Greenhouse correction was applied when sphericity is not assumed. We report
304 mean \pm SEM or median \pm SEMedian ($\frac{\text{Median absolute deviation}}{\sqrt{n}}$) for data following normal and
305 non-normal distribution, respectively.

306 **Results**

307 **Multi-photon Ca²⁺ imaging of sAM selectivity in shell IC neurons.**

308 We used a viral vector to express a genetically encoded Ca²⁺ indicator (GCaMP6f or 6s; see
309 Methods) in the IC of mice, and then used 2-photon imaging to record neural activity in shell IC
310 neurons as awake, head-fixed mice listened to sAM narrow-band noises (Figure 1A-C; 65-70 dB
311 SPL, 0.5-1 s duration, 5-200 Hz sAM rate, 0-100 % sAM depth, carrier bandwidth: 16 ± 2 kHz).
312 Imaging data were collected from fields of view (FOVs) located across the medial-lateral axis of
313 the dorsal IC 20-55 μm from the brain surface (Figure 1C).

314 We recorded a total of 1986 regions of interest (i.e., neurons) in 19 fields of view from N = 8
315 mice. Of these, n = 689 (34.7 %) were significantly sound-responsive as determined by a ‘signal
316 autocorrelation’ bootstrapping analysis (Geis et al., 2011; Wong and Borst, 2019) and were used
317 for further analysis. In agreement with classic single- and multi-unit data from central IC neurons
318 (Rees and Møller, 1983, 1987; Krishna and Semple, 2000; Nelson and Carney, 2007), we found
319 that n = 409 (59.4 %) sound-responsive shell IC neurons displayed a sound-evoked increase in
320 fluorescence, i.e., they increased firing rates to sAM sounds (Figure 1D, left). Interestingly, n =
321 280 (40.6 %) neurons were sound-inhibited and *decreased* their fluorescence during sound
322 presentation (Figure 1D, right); this fluorescence decrease likely reflects sound-evoked
323 suppression of baseline firing rates (Wong and Borst, 2019).

324 We next quantified sAM selectivity of shell IC neurons by calculating lifetime sparseness, a
325 measure of how discriminately a neuron responds to a set of stimuli over an observation period,
326 for each sound-excited and sound-inhibited neuron (Rolls and Tovee, 1995; Vinje and Gallant,
327 2000; Willmore and Tolhurst, 2001; Kato et al., 2017). Lifetime sparseness values range from 0
328 to 1, reflecting the extent to which a neuron’s sound-evoked fluorescence change is unique to a

329 single sAM sound (sparseness = 1) or alternatively, all sounds tested (sparseness = 0). Across the
330 population, sound-excited neurons showed a larger spread of lifetime sparseness values, and thus
331 greater selectivity to specific sAM rates compared to sound-inhibited neurons (Figure 1F; U =
332 182294, z = 16.051, p < 0.0001, Mann-Whitney U test). Indeed, sound-inhibited neurons were
333 often broadly inhibited by all sounds tested, whereas sound-excited neurons typically responded
334 maximally to a specific tested sAM rate (see examples in Figure 1D). Consequently, many
335 sound-excited shell IC neurons displayed the canonical modulation transfer functions previously
336 observed in the central IC (Langner and Schreiner, 1988; Krishna and Semple, 2000; Nelson and
337 Carney, 2007): Low-pass, high-pass, band-pass, and band-reject (Figure 1G-L). The peak
338 fluorescence response to preferred sAM rate generally increased as a function of sAM depth
339 (Figure 1G-K, right panels), indicating that higher depths drove stronger firing rates. Altogether
340 these data show that modulation transfer functions previously observed in the central IC are also
341 found in the shell IC, where the neuronal response to sAM sounds increases as its depth
342 increases. However, many neurons often had significant Ca²⁺ fluorescence responses even at
343 non-preferred sAM rates, in agreement with the previously reported broad and variable feature
344 selectivity of non-lemniscal IC neurons (Aitkin et al., 1975; Ito et al., 2014; Parras et al., 2017;
345 Wang and Borst, 2019; Chen and Song, 2019).

346 **GABAergic and glutamatergic neurons in shell IC exhibit similar responses to sAM sounds**
347 The IC contains two neurochemically distinct neuron classes that release either the excitatory or
348 inhibitory neurotransmitters glutamate or GABA, respectively. However, the extent to which
349 these distinct neurons show unique sound responses is less clear (Oliver et al., 1994; Ito et al.,
350 2011; Ito and Oliver, 2012; Wong and Borst, 2019; Ito, 2020). To test whether sAM tuning
351 differed between GABAergic and glutamatergic neurons, we expressed the red fluorescent

352 protein tdTomato in GABAergic neurons in a subset of mice by crossing VGAT-ires-cre and
353 Ai14 reporter mice ($N = 4$) and compared GCaMP6f responses in tdTomato-positive and -
354 negative neurons (GABAergic and presumptive glutamatergic neurons, respectively; Figure 2A,
355 B). GABAergic neurons comprise 11.5 % ($n = 57$) of the neurons in these datasets (Figure 2C).
356 In agreement with prior studies suggesting that GABAergic and glutamatergic IC neurons exhibit
357 largely overlapping sound responses (Ono et al., 2017; Wong and Borst 2019; Chen and Song
358 2019), we found that, within GABAergic neurons, 43.8 % were sound-excited and 56.2 % were
359 sound-inhibited (Figure 2D, E), the sound-excited GABAergic and glutamatergic neurons
360 showed qualitatively similar distributions of sAM selectivity and modulation transfer functions
361 (Figure 2F). These results further support a limited correlation between neurotransmitter
362 phenotype and acoustic responses in IC neurons. Given the lack of overt differences between
363 GABAergic and presumptive glutamatergic IC neurons in their responses to sAM sounds, data
364 from both cell-types were pooled for all subsequent analyses.

365 **Individual shell IC neurons transmit unreliable sAM information**

366 We further quantified sAM sensitivity by testing how shell IC neurons respond at their preferred-
367 compared to their non-preferred sAM rate. To this end, we calculated the neurometric sensitivity
368 index (d-prime, von Trapp et al., 2016) for 668 sAM preferring neurons, counting responses at
369 the preferred sAM rate as a hit, whereas responses at other sAM rates were counted as false
370 alarms (Figure 3A). Sound-excited neurons had a higher neurometric d-prime compared to
371 sound-inhibited neurons (Figure 3B; d-prime for sound-excited and inhibited neurons at 100%
372 sAM depth: 0.96 ± 0.03 and 0.20 ± 0.02 , respectively. Main effect of sAM depth: $F(4,16690) =$
373 197.6, $p < 0.0001$. Main effect of neuron type: $F(1,16690) = 912.4$, $p < 0.0001$. Interaction: $F(4,$
374 $16690) = 124.1$, $p < 0.0001$, two-way ANOVA). These results qualitatively mirror our

375 previously observed differences in lifetime sparseness (e.g., Figure 1F). When averaged across
376 the entire neuronal population, the mean d-prime increased monotonically as a function of the
377 sAM depth for the preferred sAM rate, while changing the sAM depth of non-preferred sAM
378 rates had little impact on the d-prime (Figure 3C, left). However, this effect was primarily driven
379 by $n = 89$ (13 %) sound-excited neurons whose d-prime values increased sharply as a function of
380 sAM depth (Figure 3B), and only a small population $n = 15$ (2 %) of sound-excited neurons
381 showed the opposite trend, i.e., neurometric discrimination being inversely proportional to sAM
382 depth (Figure 3C). This result may reflect the fact that in sound-excited neurons, increasing sAM
383 depth preferentially increased fluorescence responses at the neuron's preferred sAM rate,
384 whereas increasing sAM depth of non-preferred sAM rates had minimal effect on neuronal
385 activity (Figure 3D; Main effect of sAM depth: $F(1.488, 598.1) = 44.48, p < 0.0001$. Main effect
386 of preference: $F(1, 402) = 99.79, p < 0.0001$. Interaction: $F(1.708, 686.8) = 60.73, p < 0.0001$,
387 two-way repeated-measures ANOVA). Altogether, these data show that a subset of shell IC
388 neurons can reliably discriminate sAM rates, in a sAM depth-dependent manner. However, most
389 individual neurons still displayed poor d-prime profiles especially at low sAM depths (Figure
390 3B), suggesting that many shell IC neurons transmit variable, and potentially unreliable,
391 information regarding sAM features.

392 **Neural population coding of sAM rate**

393 Our data suggest that single neuron discrimination of sAM rate is variable, particularly at low
394 sAM depths. However, in circuits where the feature selectivity of individual neurons is noisy or
395 ambiguous, neural population codes (i.e., collective activity in large sets of neurons) nevertheless
396 provide an accurate representation of sensory signals for downstream circuits (Partridge et al.,

397 1981; Lee et al., 1988; Vogels, 1990; Heiligenberg, 1991; Stringer et al., 2021; Robotka et al.,
398 2023).

399 We thus tested the extent to which sAM features are encoded in the activity of shell IC neuronal
400 *populations*. While the pattern in population activity is hard to interpret mechanistically, machine
401 learning algorithms are valuable tools to extract meaningful information from neural population
402 codes. To identify neural population response pattern to sAM sounds while preserving the
403 temporal fidelity of the neural data, we employed a contemporary artificial convolutional neural
404 network (CNN), similar to previous approaches applied to visual and auditory systems
405 (Anumanchipalli et al., 2019; Sun et al., 2020; Zhang et al., 2020).

406 We trained the CNN to decode sAM features using time series fluorescence responses from all
407 significantly sound-responsive shell IC neurons in our datasets (Figure 4A). We first tested if the
408 CNN could jointly classify the sAM depth and -rate of the sound on a trial-by-trial basis.
409 Accordingly, joint decoding accuracy significantly exceeded chance level ($W = 190$, $p = 0.0001$,
410 Wilcoxon signed-rank test), and the sAM rate was classified more accurately at higher depths
411 (Figure 4B). These data suggest that the representational fidelity of sAM rate in shell IC
412 populations increases with sAM depth, such that a CNN's sAM rate classification accuracy
413 should vary with the sAM depth of its training data. To test this hypothesis, we next trained
414 CNNs on data from all sAM rates under a specific sAM depth, and quantified rate decoding
415 accuracy at each depth. Accuracy was not significantly different from chance level when CNNs
416 were trained on 20 % sAM depth (Figure 4C). However, sAM rate classification accuracy
417 increased sharply above chance level when CNNs were trained with higher sAM depths, peaking
418 at $81\% \pm 4\%$ when trained on fully modulated sAM sounds (Figure 4C, D. Main effect of sAM
419 depth: $F(2.496, 44.92) = 17.83$, $p < 0.0001$. Main effect of treatment group (trained vs. chance

420 level): $F(1,18) = 127.7$, $p < 0.0001$. Interaction: $F(2.517, 45.31) = 25.92$, $p < 0.0001$, two-way
421 repeated-measures ANOVA). Moreover, classification errors tended to cluster along the diagonal
422 of confusion matrices (Figure 4C), implying that the population representation of sAM rates
423 varies along a continuum. Importantly, classification accuracy was relatively impervious to
424 changes in hyperparameters such as the kernel size in the convolutional layer ($\chi^2(4) = 6.42$, $p =$
425 0.44, Friedman test for sAM rate decoding accuracy under 100% depth with different kernel
426 size). Moreover, shuffling neurons in population response matrices had little impact on the
427 decoding performance of sAM rate classification ($W = 42.5$, $p = 0.5571$, Wilcoxon signed-rank
428 test for sAM rate decoding accuracy under 100% depth when neuron is shuffled and control).
429 Altogether, these results suggest that sAM rate representation in the shell IC is supported by a
430 population code such that representational fidelity is steeply dependent on sAM depth. These
431 results are in qualitative agreement with perceptual thresholds determined from behaving rodents
432 engaged in AM detection tasks (Kelly et al., 2006; Caras and Sanes, 2017; van den Berg et al.,
433 2023).

434 **Noisy population codes provide reliable estimates of sAM features**

435 The CNN model's high performance may reflect a robust capacity for sAM rate discrimination
436 despite an otherwise low sensitivity at level of individual shell IC neurons. In agreement, the
437 average area under receiver operating characteristic (ROC) curve (AUC) of the sAM rate
438 classifier under 100 % sAM depth reached 0.96 (Figure 4E). We further tested the classifier's
439 discrimination ability by calculating a d-prime for sAM rate classification under 100 % depth.
440 The CNN classifier achieved a mean d-prime of 2.16 ± 0.29 , more than 3-fold higher than the
441 mean neurometric d-prime for single neurons at 100 % sAM depth (0.66 ± 0.02 ; Figure 4F).
442 Importantly, this result did not simply reflect the classifier's reliance on a minority of shell IC

443 neurons that are highly tuned to specific sAM rates: We conducted an *in silico* “lesion”
444 experiment where we trained the sAM rate decoder while excluding from training datasets highly
445 tuned neurons, as characterized by top deciles of d-prime (10 %, 20 %, and 30 %; Figure 4G).
446 We then tested the extent to which sAM rate classification accuracy deteriorates relative to
447 control training conditions. We minimized the effect of neuron count on classifier accuracy by
448 randomly excluding the same number of neurons from control classifiers as removed from the
449 “lesion” condition (Averbeck et al., 2006; Yoshida and Ohki, 2020; Zhang et al, 2020; Stringer
450 et al., 2021). Unsurprisingly, the mean decoding accuracy of the sAM rate classifier was reduced
451 when trained without highly tuned neurons, although this reduction was modest: “Lesioned”
452 classifiers nevertheless achieved a greater than three-fold greater accuracy above chance level
453 (Figure 4H, $74\% \pm 5\%$, $66\% \pm 5\%$, and $58\% \pm 5\%$ for removal of 10, 20, and 30 % top
454 deciles respectively. A two-way ANOVA revealed a main effect of treatment group (i.e. lesion
455 vs. control: $F(1, 86) = 10.91, p = 0.0014$) and a main effect of lesion size ($F(2, 86) = 3.353, p =$
456 0.0398) but no significant interaction ($F(2, 86) = 0.2909, p = 0.7483$), and Šídák’s multiple
457 comparison tests revealed no statistical significance between lesion and control groups across all
458 lesion sizes (Figure 4H). Thus, we conclude that sAM rate identity is robustly represented via a
459 population code, despite a noisy and otherwise low neurometric discrimination measured in
460 individual neurons.

461 **Sound-inhibited neurons contribute minimally to sAM population codes**

462 Our training datasets thus far included all sound-responsive neurons regardless of whether they
463 were excited or inhibited by sAM sounds. However, our results of Figure 1F show that sound-
464 inhibited neurons generally exhibit low lifetime sparseness and thereby are broadly tuned to
465 sAM rates. To what extent do sound-inhibited neurons contribute to sAM rate representations?

466 We first tested this idea by removing sound-excited neurons from the training datasets, such that
467 sAM rate classifier training progressed exclusively using sound-inhibited neurons. Interestingly,
468 classifiers trained under these conditions showed decoding accuracies that differed significantly
469 from chance-level. However, the summary data lacked the strong relationship between sAM
470 depth and decoding accuracy observed in Figure 4D, where classifiers were trained on the entire
471 neuronal population (Figure 5A; Main effect of sAM depth: $F(2.9, 46.4) = 1.023, p = 0.3891$.
472 Main effect of treatment group: $F(1, 16) = 14.36, p = 0.0016$. Interaction: $F(3.382, 54.12) =$
473 $4.725, p = 0.0039$, two-way repeated-measures ANOVA). Consequently, classifiers trained
474 exclusively on sound-inhibited neurons exceeded chance level only when trained with 60% and
475 100% depth data (Figure 5A, Šídák's multiple comparison tests), which qualitatively aligns with
476 our analyses showing mostly uniform sound-evoked inhibition in shell IC neurons (Figure 1).

477 We next tested how removing sound-inhibited neurons from the training datasets impacts sAM
478 rate decoding accuracy. To ensure that any result does not reflect a spurious effect of imbalanced
479 sound excited and inhibited neuron counts ($n = 409$ and $n = 280$, respectively; Figure 1), we
480 randomly removed subgroups of sound-excited neurons such that the total neuron number in
481 each training session was the same as for classifiers trained exclusively with sound-inhibited
482 neurons. CNNs trained on balanced sound-excited neurons recapitulated the depth-dependent
483 increase in classification accuracy observed in Figure 4D (Figure 5B; Main effect of sAM depth:
484 $F(2.637, 42.2) = 16.06, p < 0.0001$. Main effect of treatment group (trained vs. chance level): F
485 $(1, 16) = 101, p < 0.0001$. Main effect of interaction: $F(2.573, 41.16) = 20.25, p < 0.0001$, two-
486 way repeated-measures ANOVA). Consequently, sAM rate classifiers under 100 % depth trained
487 using balanced sound-excited neurons performed significantly higher than classifiers trained
488 using only sound-inhibited neurons, and had accuracies comparably to classifiers trained using

489 all sound-responsive neurons (Figure 5C; $H(2) = 27.4074$, $p < 0.0001$, Kruskal-Wallis test; See
490 Dunn's multiple comparison tests). These results indicate that activity of sound-excited neurons
491 suffices to represent sAM sound identity in the shell IC, and align with our observation that
492 sound-inhibited neurons are generally suppressed by all sAM stimuli (Figure 1F). Thus, sound-
493 evoked firing rate increases likely transmit the bulk of informative shell IC efferent signals.
494 More broadly, these conclusions agree with recent data from the primary visual cortex
495 suggesting that increases, rather than decreases in neuronal firing rate are the dominant source of
496 perceptually relevant neural activity (Cone et al., 2020).

497 **CNNs classify absolute sAM depth**

498 Detection thresholds for sAM sounds can depend on sAM rate in rodents, birds and humans
499 (Dooling and Searcy, 1981; Bacon and Viemeister, 1985; Klump and Okanoya, 1991; van der
500 Berg et al., 2023). We therefore asked if population representations of sAM depth are
501 differentially informative depending on sAM rate. To this end, we iteratively trained CNN
502 classifiers on data from all sAM depths under a specific sAM rate and tested if sAM depth
503 classification accuracy varied as a function of the training sets' sAM rate. Similar to our analysis
504 of joint rate- and depth decoding accuracy, CNNs classified the absolute sAM depth significantly
505 above chance for all training sets, although accuracy was only marginally (and not significantly)
506 dependent on sAM rate (Figure 6A, B. Main effect of sAM rate: $F(3.4, 61.21) = 2.396$, $p =$
507 0.0694. Main effect of treatment group (trained vs. chance level): $F(1, 18) = 202.8$, $p < 0.0001$.
508 Interaction: $F(2.942, 52.96) = 3.789$, $p = 0.016$, two-way repeated-measures ANOVA).
509 Moreover, confusion matrices showed that classification errors were most likely to occur
510 between neighboring sAM depths, similar to the results observed in psychometric sAM depth
511 discrimination (Figure 6B; Lee and Bacon, 1997). Our results thus show that sAM depth is

512 reasonably well represented in the shell IC, and that this discriminative ability remains robust
513 over an order of magnitude difference in sAM rate.

514 **sAM representation diverges across different sAM depths**

515 sAM rate classification accuracy depended steeply on the sAM depth of the training set and
516 saturated near 60-80 % depth (Figure 4D). These results suggest that representational fidelity
517 increases monotonically with the depth of the sAM sounds. If this hypothesis is correct, a CNN
518 classifier trained under a single sAM depth might still decode the rate of stimuli with lower or
519 higher sAM depths at above chance accuracy. However, this hypothesis further predicts that
520 decoding accuracy under these conditions should vary as a function of the depth difference
521 between training and testing datasets.

522 To this end, we trained CNN classifiers on data from all sAM rates under a single sAM depth.
523 We then measured classification accuracy on both original testing sets held back from training
524 (e.g., datasets with the same sAM depth as the training set data), as well as extended testing sets
525 consisting of data from lower or higher sAM depths not incorporated in training (Figure 7A).
526 With the exception of CNNs trained on 20 % sAM depth, classification accuracy remained above
527 chance when evaluating held-out data of the same depth (Figure 7B, see also Figure 4D).

528 $\chi^2_{20\%}(4) = 3.027, \chi^2_{40\%}(4) = 20.38, \chi^2_{60\%}(4) = 60.20, \chi^2_{80\%}(4) = 58.82, \chi^2_{100\%}(4) =$
529 $59.22; p_{20\%} = 0.5533, p_{40\%} = 0.0004, p_{60\%}, p_{80\%}, p_{100\%} < 0.0001$, Friedman test for each sub-
530 panel in B). Interestingly, CNN performance was capped by the training set's sAM depth: When
531 tested on extended datasets comprised of higher sAM depths than the training set, decoding
532 accuracy rarely exceeded the accuracy of the held-out data (Figure 7B; see results of post-hoc
533 tests). Furthermore, decoding accuracy decreased monotonically instead of randomly as a
534 function of distance between training and testing sAM depths when the extended datasets' sAM

535 depth was lower than that of training data; this effect is particularly striking when the model is
536 trained at 80 % and 100 % depths (Figure 7B). One potential interpretation is that population
537 responses to different sAM rates become increasingly dissimilar at higher sAM depths.
538 Additionally, increasing sAM depth likely drives higher firing rates in shell IC neurons, thereby
539 increasing the amplitude of Ca^{2+} signals which reflect mean firing rates in a specific time
540 window. Either of these scenarios could account for the observation that classifiers trained on
541 high depths can robustly classify sAM rate at lower depths, whereas increasing testing sAM
542 depth has little effect when the model was exclusively trained on low sAM depths (Figure 7B).

543 We further quantified the extent of similarity between different sAM depths using pattern
544 correlation (Friedrich and Laurent, 2001; Deitch et al., 2021; see Methods). The averaged
545 correlation data during sound presentation reveal an increased correlation as a function of sAM
546 depth similarity (Figure 7C, right, heatmaps). These data further argue that under passive
547 listening conditions, sAM depth appears to be represented in the shell IC population activity as a
548 continuum, rather than as distinct categories. This similarity is also reflected in the low-
549 dimensional representation of neural population data by applying a t-distributed stochastic
550 neighbor embedding (t-SNE) clustering analysis (van der Maaten and Hinton, 2008), which
551 reduces dimensionality while preserving the similarity of data in the original dimensions. The t-
552 SNE map constructed from baseline data 1 s prior to sound onset is unstructured (Figure 7D,
553 left), whereas the map generated from data points during sound presentation results in distinct
554 clusters (Figure 7D, right). Instead of forming isolated islands, the sound data are spatially
555 clustered and interconnected based on their similarity in sAM depth. Interestingly, the t-SNE
556 data points were also separated based on sAM rate (Figure 7D, note the colors). However, this
557 separation along the rate dimension was dictated by sAM depth: When performing separate t-

558 SNE analyses using data from single sAM depths, clusters representing specific sAM rates
559 increasingly diverged in a depth-dependent manner (Figure 7E).

560 As a separate test of how sAM rate population codes vary with sAM depth, we determined
561 CNNs' sAM depth thresholds by training the models to perform a binary classification task
562 reporting the absence or presence of sAM in a band-pass noise carrier (Figure 8A). Similar to our
563 results with CNNs classifying the absolute sAM depth (Figure 5), sAM detection accuracy
564 increased as a function of the sAM depth and was marginally dependent on sAM rate (Figure
565 8B-C; Main effect of sAM depth: $F(2.196, 37.33) = 30.3, p < 0.0001$. Main effect of sAM rate:
566 $F(2.814, 47.84) = 4.54, p = 0.008$. Interaction: $F(5.755, 97.83) = 2.03, p = 0.0715$, two-way
567 repeated-measures ANOVA). The somewhat lower classification accuracy at slower sAM rates
568 is not surprising, as this results qualitatively mirrors the observation that mice's sAM detection
569 thresholds are slightly increased at slower sAM rates (van den Berg et al., 2023). Additionally, a
570 logistic function fit the mean detection accuracy across all sAM rates revealed that sAM
571 detection threshold in this binary classification task is $\sim 52\%$ sAM depth (Figure 8C). Finally,
572 we also observed that the pattern correlation between unmodulated carrier and sAM sounds
573 decreased monotonically as a function of sAM depth (Figure 8D), similar to our results
574 correlating fluorescence data of different sAM depth trials (Figure 7C). Altogether, the data
575 suggest that increasing sAM depth dictates the intensity of neuronal firing rates, whereas sAM
576 rate is represented by a combinatorial code of distinct neurons firing to subsets of sAM sounds.
577 Consequently, tuning to sAM rate at the population level remains largely unchanged across
578 different sAM depths at suprathreshold levels. On the other hand, increasing sAM depth
579 sharpens the distinction between representations of different sAM rates.

580 **Minimum resolution of sAM rate representation**

581 Humans, non-human primates, and budgerigars can discriminate 2-5 % rate differences in fully
582 modulated sAM sounds, with rodents showing slightly higher difference limens (Table 1). Given
583 our data showing that fully modulated sAM sounds can be decoded with high accuracy from
584 shell IC neural population activity, we next estimated the sAM rate resolution of shell IC neuron
585 populations. We addressed this question by presenting a series of sAM sounds with narrowly
586 spaced sAM rates (100 % sAM depth, logarithmically spaced between 30-150 Hz with 5 %
587 difference between adjacent stimuli) to a separate group of mice expressing GCaMP8s in shell
588 IC neurons ($n = 520$ sound-responsive neurons in $n = 9$ non-overlapping FOVs from $N = 5$
589 mice). Individual shell IC neurons had diverse modulation transfer functions similar to our
590 experiments in Figures 1-8 in GCaMP6f/s expressing mice (Figure 9A). Sound-inhibited neurons
591 also had lower neurometric d-prime values compared to sound-excited neurons in these
592 experiments, as observed previously using 6f/6s and more widely spaced sAM rates (Figure 9B;
593 0.35 ± 0.05 and 1.29 ± 0.07 , for sound inhibited and excited neurons, respectively). However,
594 mean d-prime values were similar across the 6f/6s and 8s data sets (Main effect of neuron type: F
595 ($1, 1064$) = 16.64 , $p < 0.0001$. Main effect of dataset: $F (1, 1064) = 2.1$, $p = 0.1476$. Interaction:
596 $F (1, 1064) = 0.3125$, $p = 0.5763$, two-way ANOVA). Moreover, the spatial arrangement of sAM
597 rates data points in t-SNE map was a continuous sequence, with similar sAM rates clustering
598 nearby (Figure 9C).

599 To estimate the lower bound of sAM rate differences that can be transmitted via population
600 codes, we adjusted the CNN decoder to perform a regression task while restricting the
601 architecture to a single convolutional layer. In this paradigm, the median decoding error was 0.18
602 ± 0.01 octaves (9.12 ± 0.29 Hz; Figure 9D, E). Although few data exist regarding mouse sAM
603 rate discrimination thresholds, the sAM rate resolution in our results is comparable to that of

604 species more sensitive to sound's temporal characteristics, such as primates and budgerigars
605 (Table 1). We thus expect that sAM rate resolution of shell IC neural population could
606 approximate and may even exceed mouse discrimination thresholds for sAM rates, but further
607 studies are required to directly test this hypothesis.

608 Finally, we tested if neuronal responses to sAM rates in the shell IC population depend on carrier
609 parameters. However, sAM tuning in single neurons appeared minimally affected when testing
610 two distinct carrier parameters (8 or 16 kHz central frequency of band-limited noise; Figure 9F-
611 H). To further test this idea, we trained the CNN on neural population responses from sAM
612 sounds with either of the two noise carriers. After training, the decoder was tested on either held-
613 out data from the same carrier parameters as data in the training sets, or on extended sets of a
614 different carrier frequency as the training sets. The median decoding error was similar when
615 evaluated on held-out and extended data sets (Figure 9I-K; 0.16 ± 0.01 and 0.20 ± 0.01 octaves,
616 respectively; $W = 11031$, $p = 0.455$, Wilcoxon signed-rank test), confirming that the sAM rate
617 representations are minimally impacted by the carrier's spectral content.

618 **Discussion**

619 Using *in vivo* 2-photon Ca^{2+} imaging, we determined how sAM rate- and depth are encoded in
620 the activity of shell IC neuron populations. An important consideration is that sAM features can
621 be encoded by average firing rates, temporal synchrony of spikes relative to the sound's
622 envelope, or a mix of both (Joris et al., 2004). While the slow kinetics of Ca^{2+} signals cannot
623 resolve phase locking of spikes to the sound envelope, peak GCaMP fluorescence is a reasonably
624 good proxy for spike rates across a particular time period (Chen et al., 2013; Ranganathan et al.,
625 2018; Beaulieu-Laroche et al., 2019; Wong and Borst, 2019). Consequently, our data strictly

626 measure the spike rate selectivity of shell IC neurons to sAM features, and we cannot rule out
627 that additional information might be encoded in the temporal patterns of shell IC neuron spiking.
628 When considering this caveat, our decoding results should be considered lower bounds for
629 representational fidelity of sAM features, as any additional temporal information would
630 presumably enhance sAM representations. However, the kinetics of colliculo-thalamic synaptic
631 potentials (~50 ms halfwidth; Bartlett and Smith, 1999), as well as the membrane time constant
632 of thalamic relay neurons (~30 ms; Venkataraman and Bartlett, 2013) will likely limit the extent
633 to which any temporal information is transmitted to downstream targets. These collective results
634 further highlight the IC's role as an important early site for rate coding of time-varying sounds
635 (Dicke et al., 2007).

636 Individual shell IC neurons were generally broadly tuned to sAM sounds, leading to perhaps
637 unreliable representation of sAM features at the single neuron level. Nevertheless, instead of
638 relying on single neuron activity, information was conveyed more accurately by population
639 responses, in agreement with a recent study showing that ensemble neural discrimination of
640 vocal signals correlates poorly with single unit selectivity in the avian auditory cortex (Robotka
641 et al., 2023). Indeed, sAM rate representational fidelity remained significantly above chance
642 even after excluding individual highly tuned neurons from the neural population, suggesting that
643 the discriminative capacity of individual neurons is not necessarily a robust index of
644 population-level representations in the shell IC layers. Accordingly, population codes are
645 commonly observed in sensory and motor systems. For instance, mouse piriform cortex
646 represents odor identity based on neural population (Lurilli and Datta, 2017) and population
647 activity of V4 of macaque monkey codes the shape of visual object (Pasupathy and Connor,
648 2002). Of note, neural population codes are hypothesized to resolve highly precise information

649 sufficient to direct behavior (Partridge et al., 1981; Georgopoulos et al., 1986; Lee et al., 1988;
650 Safaai et al., 2013). Consequently, populations of broadly responsive neurons could effectively
651 transmit multi-dimensional variables that characterize other complex sounds (Zhang and
652 Sejnowski, 1999), such as the fundamental frequency of vowel-like sounds (Carney et al., 2015;
653 Carney, 2018), frequency modulated sweeps, or harmonic stacks.

654 The fidelity of sAM rate representations depended on sAM depth, with multiple analytical
655 approaches converging upon qualitatively similar estimates of “discrimination threshold” for the
656 shell IC population activity. Specifically, data points in t-SNE mappings of neural population
657 responses to less than 40% sAM depth were barely separable, whereas distinct clusters of sAM
658 rates emerged in the 60% sAM depth mapping plot (Figure 7E). This divergence was mirrored in
659 the pattern correlation analysis across different sAM depths, where there was a sudden drop in
660 pattern correlation between 100% and 40% sAM depths (Figure 7C). sAM rate classifier
661 confusion matrices also agreed with these observations: 60% and 40% sAM depth were in the
662 “watershed area” of CNNs’ decoding performance for sAM rate classification (Figure 4), with a
663 similar estimated sAM *detection* threshold from a binary classification task (~ 52 % sAM depth;
664 Figure 8). Although future studies will be required to directly test the extent to which shell IC
665 neurons causally contribute to sAM detection or discrimination, the sAM depth thresholds
666 estimates from our neurophysiological data are in qualitative agreement with recently reported
667 behavioral detection thresholds in mice (20-30 % sAM depth; van den Berg et al., 2023).
668 Intriguingly, van den Berg et al. (2023) observed that a higher depth is required for mice to
669 detect sAM sounds as its rate decreases, and this dependence on sAM rate is also mirrored in our
670 binary classification results.

671 Non-lemniscal thalamic targets of shell IC neurons integrate heterogeneous signals (Lesicko et
672 al., 2020; Liu et al., 2023; Ibrahim et al., 2023), and disynaptically relay this information to
673 limbic circuits that orchestrate learned and innate behaviors (Miura et al., 2020; Carcea et al.,
674 2021; Valtcheva et al., 2023). In contrast to a temporal code, the rate-based encoding of time-
675 varying sounds might allow for a more efficient integration with multisensory, motor, and
676 somatosensory signals across a specific temporal window. Importantly, the temporal
677 characteristics of the sound envelope conveys information to discriminate conspecific
678 vocalizations, and many second-order targets of shell IC neurons are strongly selective for
679 species-specific vocalizations (Gadziola et al., 2012; Grimsley et al., 2013; Gadziola et al., 2016,
680 Hamilton et al., 2021). However, this selectivity may not reflect computations inherent to limbic
681 circuits, but rather could arise upstream in the midbrain: Aitkin et al., (1994) showed that 75 %
682 of feline shell IC neurons respond stronger to vocal stimuli than noise or characteristic frequency
683 stimuli while only ~25 % of central nucleus neurons do. In mice, many neurons in the low
684 frequency (presumably dorsal) IC regions respond strongly to vocalizations (Portfors et al.,
685 2009). AM features significantly shape the vocalization selectivity in some of these neurons
686 (Holmstrom et al., 2010), and prominent selectivity to ethologically relevant signals is also
687 observed in the shell IC of bats, rats and gerbils (Holmstrom et al., 2007; Gao et al., 2015;
688 Lawlor et al., 2023). In tandem with our current results, these data imply that a shell IC neural
689 population play important roles in processing and integrating vocal signals, thereby shedding
690 light on the perceptual building blocks of behavioral responses in vocal communication.

691 **References**

- 692 Ayala YA, Udeh A, Dutta K, Bishop D, Malmierca MS, Oliver DL (2015) Differences in the
693 strength of cortical and brainstem inputs to SSA and non-SSA neurons in the inferior
694 colliculus. *Sci Rep* 5:10383.
- 695 Aitkin L, Tran L, Syka J (1994) The responses of neurons in subdivisions of the inferior
696 colliculus of cats to tonal, noise and vocal stimuli. *Exp Brain Res* 98:53-64.
- 697 Aitkin LM, Webster WR, Veale JL, Crosby DC (1975) Inferior colliculus. I. Comparison of
698 response properties of neurons in central, pericentral, and external nuclei of adult cat. *J*
699 *Neurophysiol* 38:1196-1207.
- 700 Anumanchipalli GK, Chartier J, Chang EF (2019) Speech synthesis from neural decoding of
701 spoken sentences. *Nature* 568:493-498.
- 702 Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and
703 computation. *Nat Rev Neurosci* 7:358-66.
- 704 Bacon SP, Viemeister NF (1985) Temporal modulation transfer functions in normal-hearing and
705 hearing-impaired listeners. *Audiology* 24:117-134.
- 706 Bagur S, Bourg J, Kempf A, Tarpin T, Bergaoui K, Guo Y, Ceballo S, Schwenkgrub J, Verdier
707 A, Puel JL, Bourien J, Bathellier B (2023) A spatial code for temporal cues is necessary
708 for sensory learning. *bioRxiv* 520391. doi: <https://doi.org/10.1101/2022.12.14.520391>.
- 709 Barnstedt O, Keating P, Weissenberger Y, King AJ, Dahmen JC (2015) Functional
710 Microarchitecture of the Mouse Dorsal Inferior Colliculus Revealed through In Vivo
711 Two-Photon Calcium Imaging. *J Neurosci* 35:10927-10939.
- 712 Bartlett EL, Smith PH (1999) Anatomic, intrinsic, and synaptic properties of dorsal and ventral
713 division neurons in rat medial geniculate body. *J Neurophysiol* 81:1999-2016.

- 714 Beaulieu-Laroche L, Toloza EHS, Brown NJ, Harnett MT (2019) Widespread and Highly
715 Correlated Somato-dendritic Activity in Cortical Layer 5 Neurons. *Neuron* 103:235-
716 241.e4.
- 717 Bendor D, Wang X (2007) Differential neural coding of acoustic flutter within primate auditory
718 cortex. *Nat Neurosci* 10:763-771.
- 719 Bordi F, LeDoux JE (1994) Response properties of single units in areas of rat auditory thalamus
720 that project to the amygdala. II. Cells receiving convergent auditory and somatosensory
721 inputs and cells antidromically activated by amygdala stimulation. *Exp Brain Res* 98:275-
722 286.
- 723 Cai D, Yue Y, Su X, Liu M, Wang Y, You L, Xie F, Deng F, Chen F, Luo M (2019) Distinct
724 anatomical connectivity patterns differentiate subdivisions of the nonlemniscal auditory
725 thalamus in mice. *Cereb Cortex* 29:2437-2454.
- 726 Carney LH (2018) Supra-Threshold Hearing and Fluctuation Profiles: Implications for
727 Sensorineural and Hidden Hearing Loss. *J Assoc Res Otolaryngol* 19:331-352.
- 728 Carney LH, Li T, McDonough JM (2015) Speech Coding in the Brain: Representation of Vowel
729 Formants by Midbrain Neurons Tuned to Sound Fluctuations. *eNeuro* 2:ENEURO.0004-
730 15.2015.
- 731 Carcea I, Caraballo NL, Marlin BJ, Ooyama R, Riceberg JS, Mendoza Navarro JM, Opendak M,
732 Diaz VE, Schuster L, Alvarado Torres MI, Lethin H, Ramos D, Minder J, Mendoza SL,
733 Bair-Marshall CJ, Samadjopoulos GH, Hidema S, Falkner A, Lin D, Mar A, Wadghiri
734 YZ, Nishimori K, Kikusui T, Mogi K, Sullivan RM, Froemke RC (2021) Oxytocin
735 neurons enable social transmission of maternal behaviour. *Nature* 596:553-557.

- 736 Caras ML, Sanes DH (2017). Top-down modulation of sensory cortex gates perceptual
737 learning. PNAS 114:9972-9977.
- 738 Champoux F, Paiement P, Mercier C, Lepore F, Lassonde M, Gagné JP (2007) Auditory
739 processing in a patient with a unilateral lesion of the inferior colliculus. Eur J Neurosci
740 25:291-297.
- 741 Chen C, Cheng M, Ito T, Song S (2018) Neuronal Organization in the Inferior Colliculus
742 Revisited with Cell-Type-Dependent Monosynaptic Tracing. J Neurosci 38:3318-3332.
- 743 Chen C, Song S (2019) Differential cell-type dependent brain state modulations of sensory
744 representations in the non-lemniscal mouse inferior colliculus. Commun Biol 2:356.
- 745 Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA,
746 Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive
747 fluorescent proteins for imaging neuronal activity. Nature 499:295-300.
- 748 Cone JJ, Bade ML, Masse NY, Page EA, Freedman DJ, Maunsell JHR (2020) Mice
749 Preferentially Use Increases in Cerebral Cortex Spiking to Detect Changes in Visual
750 Stimuli. J Neurosci 40:7902-7920.
- 751 Cortes C, Mohri M, Rostamizadeh A (2012) L2 Regularization for Learning Kernels. Arxiv.
752 <https://doi.org/10.48550/arXiv.1205.2653>
- 753 Deitch D, Rubin A, Ziv Y (2021) Representational drift in the mouse visual cortex. Curr Biol
754 31:4327-4339.
- 755 Dicke U, Ewert SD, Dau T, Kollmeier B (2007) A neural circuit transforming temporal
756 periodicity information into a rate-based representation in the mammalian auditory
757 system. J Acoust Soc Am 121:310-326.

- 758 Dooling RJ, Searcy MH (1981) Amplitude modulation thresholds for the parakeet (*Melopsittacus*
759 *undulatus*). *J Comp Physiol* 143:383-388.
- 760 Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech
761 reception. *J Acoust Soc Am* 95:1053-64
- 762 Elliott TM, Theunissen FE (2009) The Modulation Transfer Function for Speech Intelligibility.
763 *PLoS Comput Biol* 5:e1000302.
- 764 Edeline JM, Weinberger NM (1992) Associative retuning in the thalamic source of input to the
765 amygdala and auditory cortex: receptive field plasticity in the medial division of the
766 medial geniculate body. *Behav Neurosci* 106:81-105.
- 767 Formby C (1985) Differential sensitivity to tonal frequency and to the rate of amplitude
768 modulation of broadband noise by normally hearing listeners. *J Acoust Soc Am* 78:70-77.
- 769 Friedrich RW, Laurent G (2001) Dynamic optimization of odor representations by slow temporal
770 patterning of mitral cell activity. *Science* 291:889-894.
- 771 Frisina RD, Smith RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil
772 cochlear nucleus: I. A hierarchy of enhancement. *Hear Res* 44:99-122.
- 773 Gadziola MA, Grimsley JM, Shanbhag SJ, Wenstrup JJ (2012) A novel coding mechanism for
774 social vocalizations in the lateral amygdala. *J Neurophysiol* 107:1047-1057.
- 775 Gadziola MA, Shanbhag SJ, Wenstrup JJ (2016) Two distinct representations of social
776 vocalizations in the basolateral amygdala. *J Neurophysiol* 115:868-886.
- 777 Gallun F, Souza P (2008) Exploring the role of the modulation spectrum in phoneme recognition.
778 *Ear Hear* 29:800-813.

- 779 Gao PP, Zhang JW, Fan SJ, Sanes DH, Wu EX (2015) Auditory midbrain processing is
780 differentially modulated by auditory and visual cortices: An auditory fMRI study.
781 Neuroimage 123:22-32.
- 782 Geis HR, Borst JG (2009) Intracellular responses of neurons in the mouse inferior colliculus to
783 sinusoidal amplitude-modulated tones. J Neurophysiol 101:2002-2016.
- 784 Geis HR, van der Heijden M, Borst JG (2011) Subcortical input heterogeneity in the mouse
785 inferior colliculus. J Physiol 589:3955-3967.
- 786 Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement
787 direction. Science 233:1416-1419.
- 788 Grimsley JM, Hazlett EG, Wenstrup JJ (2013) Coding the meaning of sounds: contextual
789 modulation of auditory responses in the basolateral amygdala. J Neurosci. 33:17538-48.
- 790 Hamilton LS, Oganian Y, Hall J, Chang EF (2021) Parallel and distributed encoding of speech
791 across human auditory cortex. Cell 184:4626-39.
- 792 Hewitt MJ, Meddis R (1994) A computer model of amplitude-modulation sensitivity of single
793 units in the inferior colliculus. J Acoust Soc Am 95:2145-2159.
- 794 Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving
795 neural networks by preventing co-adaptation of feature detectors. Arxiv.
796 <https://arxiv.org/abs/1207.0580>
- 797 Holmstrom LA, Eeuwes LB, Roberts PD, Portfors CV (2010) Efficient encoding of vocalizations
798 in the auditory midbrain. J Neurosci 30:802-819.
- 799 Holmstrom L, Roberts PD, Portfors CV (2007) Responses to social vocalizations in the inferior
800 colliculus of the mustached bat are influenced by secondary tuning curves. J
801 Neurophysiol 98:3461-3472.

- 802 Ibrahim BA, Shinagawa Y, Douglas A, Xiao G, Asilador AR, Llano DA (2023) Microprism-
803 based two-photon imaging of the lateral cortex of the mouse inferior colliculus reveals
804 novel organizational principles of the auditory midbrain. bioRxiv 515308.
805 <https://doi.org/10.1101/2022.11.05.515308>
- 806 Ito T, Bishop DC, Oliver DL (2011) Expression of glutamate and inhibitory amino acid vesicular
807 transporters in the rodent auditory brainstem. J Comp Neurol 519:316-340.
- 808 Ito T, Oliver DL (2012) The basic circuit of the IC: tectothalamic neurons with different patterns
809 of synaptic organization send different messages to the thalamus. Front Neural Circuits
810 6:48.
- 811 Ito T, Hirose J, Murase K, Ikeda H (2014) Determining auditory-evoked activities from multiple
812 cells in layer 1 of the dorsal cortex of the inferior colliculus of mice by in vivo calcium
813 imaging. Brain Res 1590:45-55.
- 814 Ito T, Bai J, Ostry DJ (2020) Contribution of sensory memory to speech motor learning. J
815 Neurophysiol 124:1103-1109.
- 816 Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds.
817 Physiol Rev 84:541-577.
- 818 Kato HK, Asinof SK, Isaacson JS (2017) Network-Level Control of Frequency Tuning in
819 Auditory Cortex. Neuron 95:412-423.e4.
- 820 Kelly JB, Cooke JE, Gilbride PC, Mitchell C, Zhang H (2006) Behavioral limits of auditory
821 temporal resolution in the rat: Amplitude modulation and duration discrimination. J
822 Comp Psychol 120:98-105.
- 823 Kingma DP, Ba J (2014) Adam: A Method for Stochastic Optimization. Arxiv.
824 <https://arxiv.org/abs/1412.6980>

- 825 Klump GM, Okanoya K (1991) Temporal modulation transfer functions in the European starling
826 (Sturnus vulgaris): I. Psychophysical modulation detection thresholds. Hear Res 52:1-11.
- 827 Koumura T, Terashima H, Furukawa S (2019) Cascaded Tuning to Amplitude Modulation for
828 Natural Sound Recognition. J Neurosci 39:5517-5533.
- 829 Koumura T, Terashima H, Furukawa S (2023) Human-Like Modulation Sensitivity Emerging
830 through Optimization to Natural Sound Recognition. J Neurosci 43:3876-3894.
- 831 Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally
832 amplitude-modulated tones in the inferior colliculus. J Neurophysiol 84:255-273.
- 833 Kurt S, Ehret G (2010) Auditory discrimination learning and knowledge transfer in mice depends
834 on task difficulty. PNAS 107:8481-8485.
- 835 Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I.
836 Neuronal mechanisms. J Neurophysiol 60:1799-1822.
- 837 Lawlor J, Wohlgemuth M, Moss CF, Kuchibhotla KV (2023) Spatially clustered neurons encode
838 vocalization categories in the bat midbrain. bioRxiv 545029
839 <https://doi.org/10.1101/2023.06.14.545029>
- 840 LeDoux JE, Farb CR, Romanski LM (1991) Overlapping projections to the amygdala and
841 striatum from auditory processing areas of the thalamus and cortex. Neurosci Lett
842 134:139-144.
- 843 Lee C, Rohrer WH, Sparks DL (1988) Population coding of saccadic eye movements by neurons
844 in the superior colliculus. Nature 332:357-360.
- 845 Lee CC, Winer JA (2008) Connections of cat auditory cortex: I. Thalamocortical system. J Comp
846 Neurol 507:1879-1900.

- 847 Lee J, Bacon SP (1997) Amplitude modulation depth discrimination of a sinusoidal carrier:
848 effect of stimulus duration. *J Acoust Soc Am* 101:3688-3693.

849 Leppla CA, Keyes LR, Glober G, Matthews GA, Batra K, Jay M, Feng Y, Chen HS, Mills F,
850 Delahanty J, Olson JM, Nieh EH, Namburi P, Wildes C, Wichmann R, Beyeler A,
851 Kimchi EY, Tye KM. Thalamus sends information about arousal but not valence to the
852 amygdala. *Psychopharmacology* 240:477-499.

853 Lesicko AMH, Sons SK, Llano DA (2020) Circuit Mechanisms Underlying the Segregation and
854 Integration of Parallel Processing Streams in the Inferior Colliculus. *J Neurosci* 40:6328-
855 6344.

856 Liu M, Xie F, Dai J, Zhang J, Yuan K, Wang N (2023) Brain-wide inputs to the non-lemniscal
857 inferior colliculus in mice. *Neurosci Lett* 793:136976.

858 Lumani A, Zhang H (2010) Responses of neurons in the rat's dorsal cortex of the inferior
859 colliculus to monaural tone bursts. *Brain Res* 1351:115-129.

860 Long GR, Clark WW (1984) Detection of frequency and rate modulation by the chinchilla. *J
861 Acoust Soc Am* 75:1184-1190.

862 Lorenzi C, Micheyl C, Berthommier F (1995) Neuronal correlates of perceptual amplitude-
863 modulation detection. *Hear Res* 90:219-227.

864 Lurilli G, Datta SR (2017) Population Coding in an Innately Relevant Olfactory Area. *Neuron*
865 93:1180-1197.e7.

866 Mackey CA, Hauser S, Schoenhaut AM, Temghare N, Ramachandran R (2022) Psychometric
867 and subcortical neurometric measures of temporal discrimination in rhesus macaques.

868 bioRxiv 502987. <https://doi.org/10.1101/2022.08.05.502987>

- 869 Mellott JG, Foster NL, Ohl AP, Schofield BR (2014) Excitatory and inhibitory projections in
870 parallel pathways from the inferior colliculus to the auditory thalamus. *Front Neuroanat*
871 8:124.
- 872 Miura I, Sato M, Overton ETN, Kunori N, Nakai J, Kawamata T, Nakai N, Takumi T (2020)
873 Encoding of social exploration by neural ensembles in the insular cortex. *PLoS Biol*
874 18:e3000584.
- 875 Moody DB (1994) Detection and discrimination of amplitude-modulated signals by macaque
876 monkeys. *J Acoust Soc Am* 95:3499-3510.
- 877 Nelson PC, Carney LH (2007) Neural Rate and Timing Cues for Detection and Discrimination of
878 Amplitude-Modulated Tones in the Awake Rabbit Inferior Colliculus. *J Neurophysiol*
879 97:522-539.
- 880 Oberle HM, Ford AN, Dileepkumar D, Czarny J, Apostolidis PF (2022) Synaptic mechanisms of
881 top-down control in the non-lemniscal inferior colliculus. *ELife* 10:e72730.
- 882 Olds J, Disterhoft JF, Segal M, Kornblith CL, Hirsh R (1972) Learning centers of rat brain
883 mapped by measuring latencies of conditioned unit responses. *J Neurophysiol* 35:202-
884 219.
- 885 Oliver DL, Morest DK (1984) The central nucleus of the inferior colliculus in the cat. *J Comp*
886 *Neurol* 222:237-264.
- 887 Oliver DL, Winer JA, Beckius GE, Saint Marie RL (1994) Morphology of GABAergic neurons
888 in the inferior colliculus of the cat. *J Comp Neurol* 340:27-42.
- 889 Ono M, Bishop DC, Oliver DL (2017) Identified GABAergic and Glutamatergic Neurons in the
890 Mouse Inferior Colliculus Share Similar Response Properties. *J Neurosci*, 37:8952–8964.

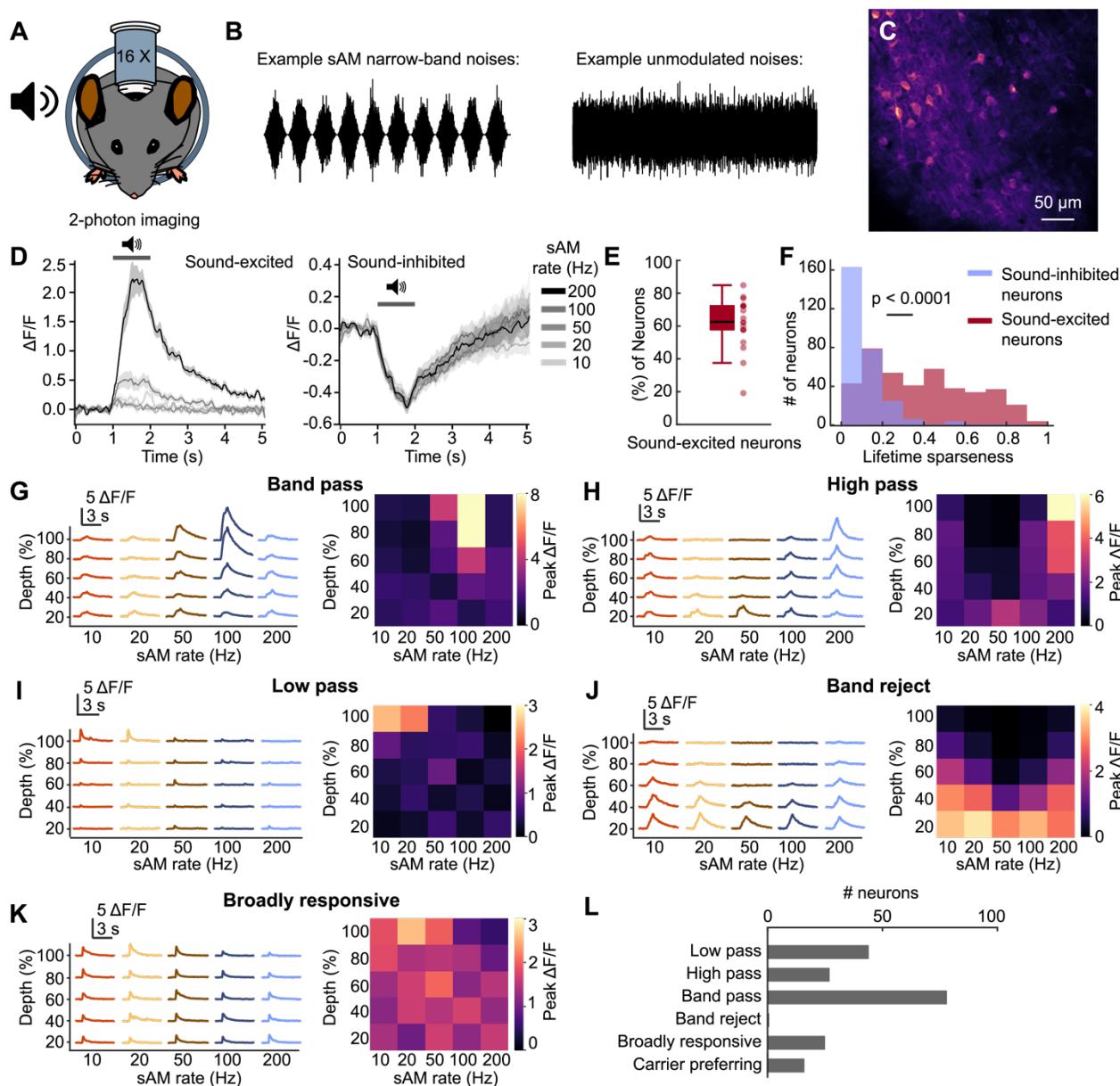
- 891 Pachitariu M, Stringer C, Dipoppa M, Schröder S, Rossi LF, Dalgleish H, Carandini M, Harris
892 KD (2016) Suite2p: beyond 10,000 neurons with standard two-photon microscopy.
893 bioRxiv 061507. <https://doi.org/10.1101/061507>
- 894 Parras GG, Nieto-Diego J, Carbajal GV, Valdés-Baizabal C, Escera C, Malmierca MS (2017)
895 Neurons along the auditory pathway exhibit a hierarchical organization of prediction
896 error. Nat Commun 8:2148.
- 897 Partridge BL, Heiligenberg W, and Matsubara J (1981) The neural basis of a sensory filter in the
898 Jamming Avoidance Response: No grandmother cells in sight. J Comp Physiol 145:153-
899 168.
- 900 Pasupathy A, Connor CE (2002). Population coding of shape in area V4. Nat Neurosci 5:1332-
901 1338.
- 902 Ponvert ND, Jaramillo S (2019) Auditory Thalamostriatal and Corticostriatal Pathways Convey
903 Complementary Information about Sound Features. J Neurosci 39:271-280.
- 904 Poremba A, Gabriel M (2001) Amygdalar efferents initiate auditory thalamic discriminative
905 training-induced neuronal activity. J Neurosci 21:270-278.
- 906 Portfors CV, Roberts PD, Jonson K (2009) Over-representation of species-specific vocalizations
907 in the awake mouse inferior colliculus. Neuroscience 162:486-500.
- 908 Ranganathan GN, Apostolides PF, Harnett MT, Xu NL, Druckmann S, Magee JC (2018) Active
909 dendritic integration and mixed neocortical network representations during an adaptive
910 sensing behavior. Nat Neurosci 21:1583-1590.
- 911 Rees A, Møller AR (1983) Responses of neurons in the inferior colliculus of the rat to AM and
912 FM tones. Hear Res 10:301-330.

- 913 Rees A, Møller AR (1987) Stimulus properties influencing the responses of inferior colliculus
914 neurons to amplitude-modulated sounds. *Hear Res* 27:129-143.
- 915 Rhode WS, Greenberg S (1994) Encoding of amplitude modulation in the cochlear nucleus of
916 the cat. *J Neurophysiol* 71:1797-825.
- 917 Robotka H, Thomas L, Yu K, Wood W, Elie JE, Gahr M, Theunissen FE (2023). Sparse
918 ensemble neural code for a complete vocal repertoire. *Cell Rep* 42:112034.
- 919 Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate
920 temporal visual cortex. *J Neurophysiol* 73:713-726.
- 921 Safaai H, von Heimendahl M, Sorando JM, Diamond ME, Maravall M (2013) Coordinated
922 population activity underlying texture discrimination in rat barrel cortex. *J Neurosci*
923 33:5843-5855.
- 924 Schiavo JK, Valtcheva S, Bair-Marshall, CJ, Song SC, Martin KA, Froemke RC (2020) Innate
925 and plastic mechanisms for maternal behaviour in auditory cortex. *Nature* 587:426-431.
- 926 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with
927 primarily temporal cues. *Science* 270:303-304.
- 928 Stringer C, Michaelos M, Tsyboulski D, Lindo SE, Pachitariu M (2021) High-precision coding
929 in visual cortex. *Cell* 184:2767-2778.
- 930 Sun P, Anumanchipalli GK, Chang EF (2020) Brain2Char: a deep architecture for decoding text
931 from brain recordings. *J Neural Eng* 17.
- 932 Syka J, Popela'r J, Kvasna'k E, Astl J (2000) Response properties of neurons in the central
933 nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. *Exp Brain
934 Res* 133:254-266.

- 935 Tan ML, Borst JG (2007) Comparison of responses of neurons in the mouse inferior colliculus to
936 current injections, tones of different durations, and sinusoidal amplitude-modulated tones.
937 J Neurophysiol 98:454-66.
- 938 Ter-Mikaelian M, Sanes DH, Semple MN (2007) Transformation of temporal properties between
939 auditory midbrain and cortex in the awake Mongolian gerbil. J Neurosci 27:6091-6102.
- 940 Valtcheva S, Issa HA, Bair-Marshall CJ, Martin KA, Jung K, Buzsáki G, Kwon HB, Froemke
941 RC (2023) Neural circuitry for maternal oxytocin release induced by infant cries. bioRxiv
942 436883. doi: <https://doi.org/10.1101/2021.03.25.436883>
- 943 van den Berg MM, Busscher E, Borst JGG, Wong AB (2023) Neurometric correlates of sensitive
944 high-frequency sound amplitude modulation detection by mice. bioRxiv 466979. doi:
945 <https://doi.org/10.1101/2021.11.02.466979>
- 946 van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. JMLR 9:2579-2605
- 947 Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during
948 natural vision. Science 287:1273-1276.
- 949 Vogels R (1990) Population coding of stimulus orientation by striate cortical cells. Biol Cybern
950 64:25-31
- 951 von Trapp G, Buran BN, Sen K, Semple MN, Sanes DH (2016) A Decline in Response
952 Variability Improves Neural Signal Detection during Auditory Task Performance. J
953 Neurosci 36:11097-11106.
- 954 Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory
955 thalamus and cortex. Neuroscience 154:294-303.

- 956 Wang X, Zhang Y, Bai S, Qi R, Sun H, Li R, Zhu L, Cao X, Jia G, Li X, Gao L (2022)
- 957 Corticofugal modulation of temporal and rate representations in the inferior colliculus of
- 958 the awake marmoset, *Cerebr Cortex*, 32:4080-4097
- 959 Willmore B, Tolhurst DJ (2001) Characterizing the sparseness of neural codes. *Network* 12:255-
- 960 270.
- 961 Winer JA, Larue DT, Diehl JJ, Hefti BJ (1998) Auditory cortical projections to the cat inferior
- 962 colliculus. *J Comp Neurol* 400:147-174.
- 963 Wong AB, Borst JG (2019) Tonotopic and non-auditory organization of the mouse dorsal
- 964 inferior colliculus revealed by two-photon imaging. *ELife* 8:e49091.
- 965 Xiong XR, Liang F, Zingg B, Ji X, Ibrahim LA, Tao HW, Zhang LI (2015). Auditory cortex
- 966 controls sound-driven innate defense behaviour through corticofugal projections to
- 967 inferior colliculus. *Nat Commun* 6:7224.
- 968 Yoshida T, Ohki K (2020) Natural images are reliably represented by sparse and variable
- 969 populations of neurons in visual cortex. *Nat Commun* 11:872.
- 970 Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K (2005)
- 971 Speech recognition with amplitude and frequency modulations. *PNAS* 102:2293-2298.
- 972 Zhang K, Sejnowski TJ (1999) Neuronal Tuning: To Sharpen or Broaden? *Neural Comput*
- 973 11:75-84.
- 974 Zhang Y, Jia S, Zheng Y, Yu Z, Tian Y, Huang T, Liu JK (2020) Reconstruction of Natural
- 975 Visual Scenes from Neural Spikes with Deep Neural Networks. *Neural networks* 125:19-
- 976 30.
- 977 Zhao HB, Liang ZA (1995) Processing of modulation frequency in the dorsal cochlear nucleus of
- 978 the guinea pig: amplitude modulated tones. *Hear Res* 82:244-256.

979 **Figures and tables:**

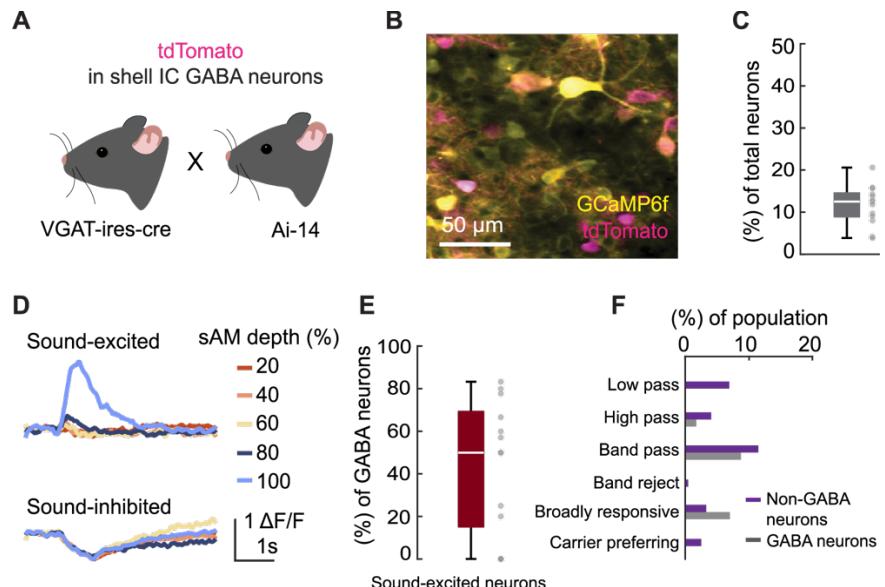


980

981 **Figure 1**

- 982 Responses of mouse shell IC neurons to sAM and unmodulated narrow-band noises. **A.**
 983 Experimental setup of sound presentation and 2-photon imaging of head-fixed, awake mice. **B.**
 984 Left, an example of presented sAM narrow-band noises (100% sAM depth and 10 Hz sAM rate).
 985 Right, an example of presented unmodulated noises. **C.** Example imaging FOV. **D.** Example

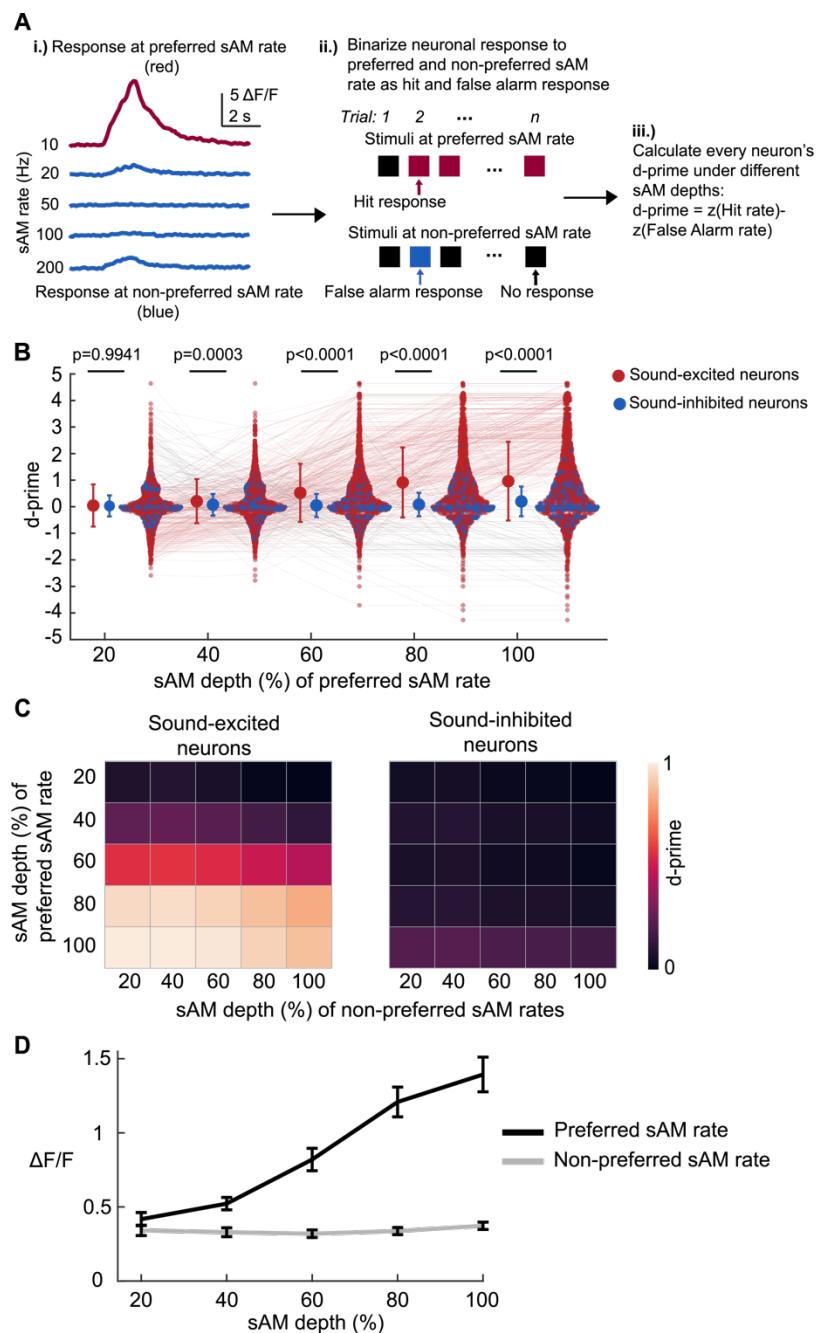
986 responses of sound-excited (left), and sound-inhibited (right) neurons to fully modulated sAM
987 sounds. Data are mean \pm SEM. **E**. Proportion of sound-excited neurons in each imaging session.
988 **F**. Distribution of lifetime sparseness of sound-excited and sound-inhibited neurons. A neuron is
989 maximally selective to sAM stimuli if its sparseness is 1 and is totally unselective if the
990 sparseness is 0. Mann-Whitney U test. **G-K**. Example of trial-averaged neuronal responses (left)
991 and peak $\Delta F/F$ (right) of band pass (**G**), high pass (**H**), low pass (**I**), and band reject (**J**), broadly
992 responsive (**K**) neurons under different sAM sounds. **L**. Distribution of the number of tuning
993 characteristics of shell IC neurons.



994

995 **Figure 2**

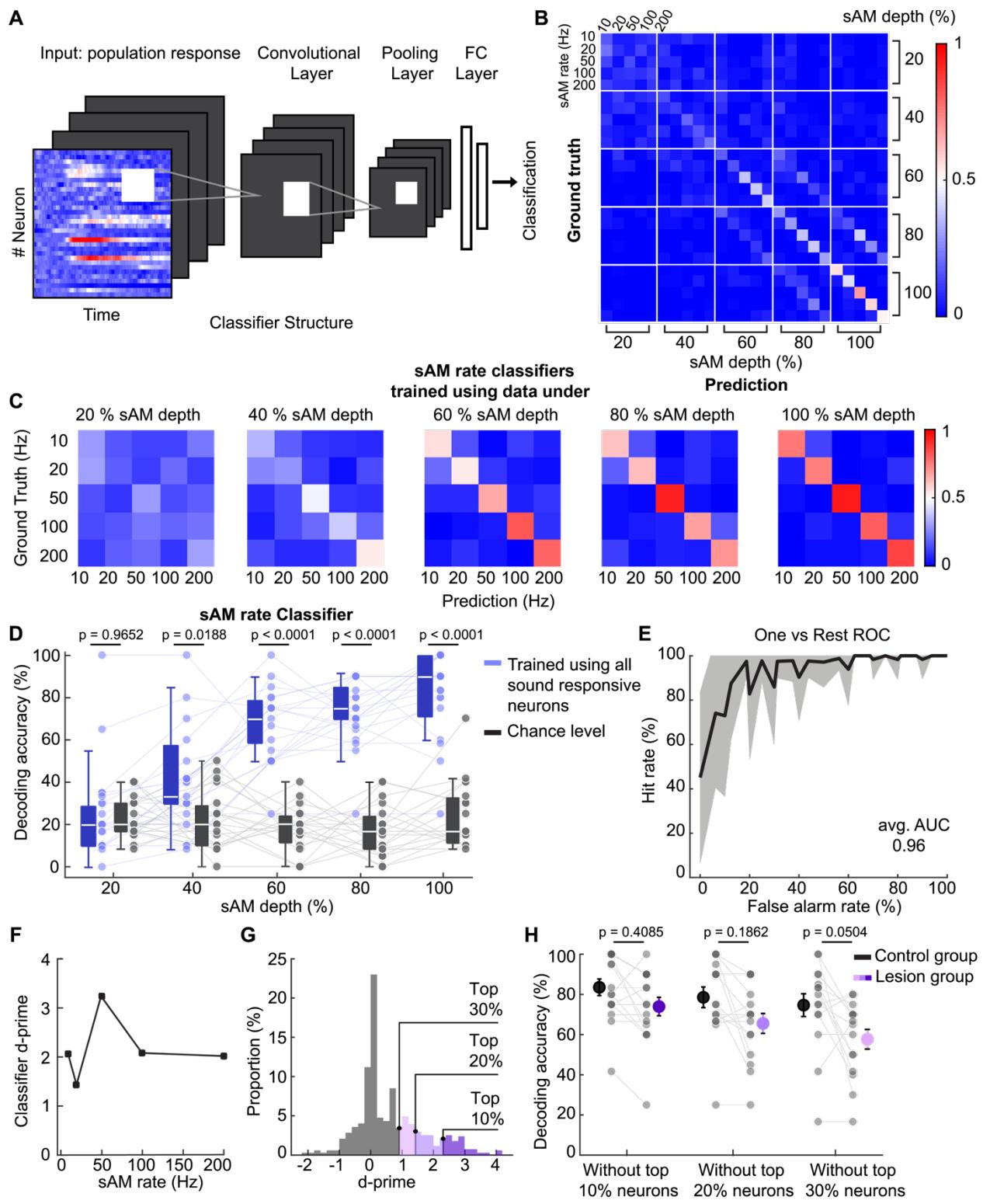
996 sAM tuning of shell IC GABAergic and non-GABAergic neurons. **A.** tdTomato was expressed
997 in GABAergic neurons in transgenic VGAT-ires-cre x Ai14 mice. **B.** Example imaging FOV. **C.**
998 Proportion of GABAergic neurons in each imaging session. **D.** Examples of sound-excited and
999 sound-inhibited responses of GABAergic neurons to sounds with 100 Hz sAM rate and different
1000 sAM depths. **E.** Proportion of sound-excited GABAergic neurons in each imaging session. **F.**
1001 Distribution of tuning characteristics of non-GABAergic and GABAergic neurons.



1003 **Figure 3**

1004 Neurometric sensitivity of individual shell IC neurons. **A.** Schematic of neurometric sensitivity
1005 index (d-prime) analysis. Preferred sAM rate is determined by the peak fluorescence response at
1006 different sAM rates (i). We binarize a neuron's response as a hit if the mean neuronal response at
1007 its preferred sAM rate exceeds three times the standard deviation of the baseline fluctuation on a

1008 single trial. Similarly, we count a false alarm response if a neuron's average response at its non-
1009 preferred sAM rate exceeds three times the standard deviation of the baseline fluctuation (ii).
1010 The d-prime was then calculated for each neuron for all pairs of sAM depths for preferred and
1011 non-preferred sAM rates (iii). **B.** Distribution of d-prime at varying sAM depths for the preferred
1012 sAM rate. Each line indicates a neuron displaying an increasing (pink) or decreasing (gray) trend
1013 across sAM depths for the preferred sAM rate (See Methods). Data are mean with \pm std. Two-
1014 way ANOVA. Šídák's multiple comparison between d-primes of sound-excited and sound-
1015 inhibited neurons. **C.** Left, averaged d-prime of sound-excited neurons. Right, averaged d-prime
1016 of sound-inhibited neurons. **D.** Trial-averaged fluorescence response to preferred and non-
1017 preferred sAM rate. Data are mean with \pm SEM.



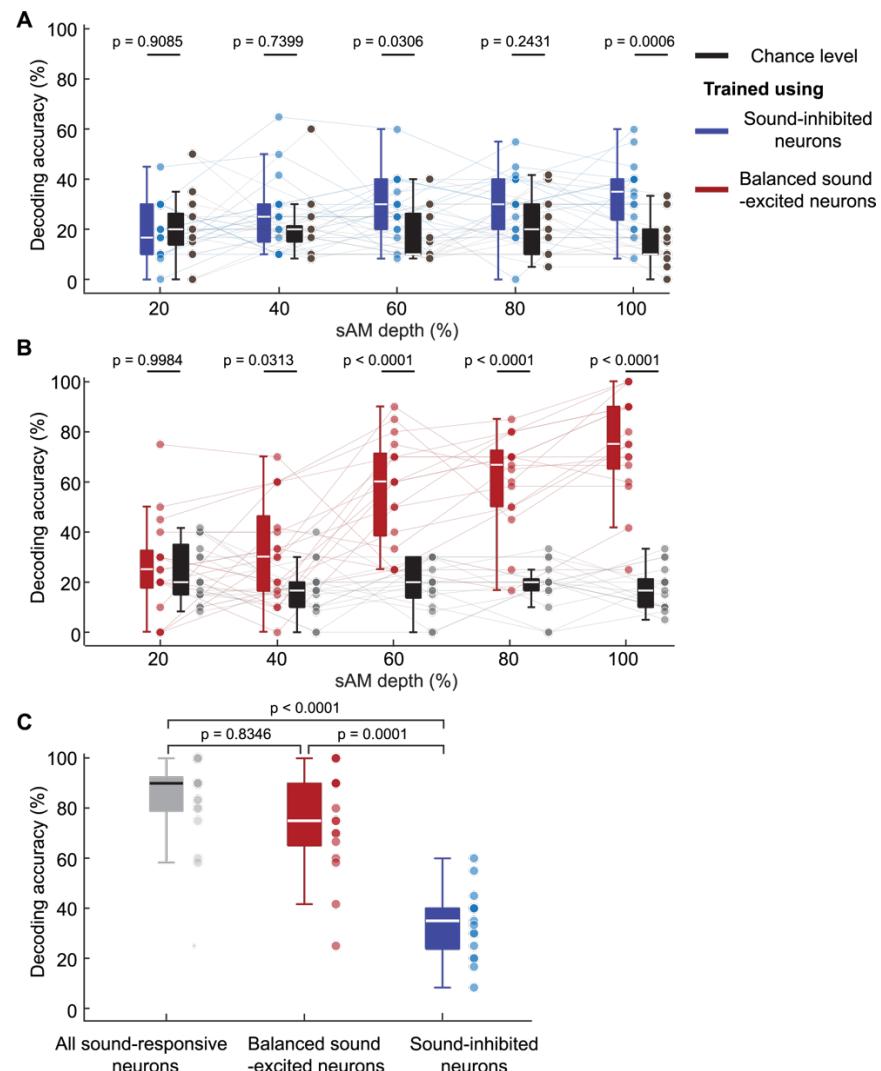
1018

1019

1020 **Figure 4**

1021 Decoding sAM features using shell IC neural population activity. **A.** Structure of the CNN
1022 classifier. A CNN classifier fed with Ca^{2+} signal time series was trained to classify sAM features.
1023 **B.** Normalized confusion matrix of sAM depth and -rate joint combination classification
1024 averaged across imaging sessions. **C.** Normalized confusion matrix of sAM rate classification
1025 under a given sAM depth averaged across imaging sessions. **D.** Decoding accuracy of the sAM
1026 rate classifier under a given depth and the corresponding chance level. Two-way repeated-
1027 measures ANOVA. Šídák's multiple comparison between the chance level accuracy and
1028 decoding accuracy trained using all neurons. **E.** One-vs-rest ROC of sAM rate classification
1029 under 100% sAM depth. Data are mean with \pm std. **F.** d-prime of sAM rate classifier under 100%
1030 sAM depth. **G.** Individual neuron d-prime distribution (under 100 % sAM depth), with the
1031 corresponding highly sensitive neurons falling within top deciles (10 %, 20 %, 30 % of d-prime).
1032 **H.** Decoding accuracy of the sAM rate classifier under 100% sAM depth, trained while
1033 excluding the top deciles (10%, 20%, 30%) of neurons, selected based on their d-prime profiles
1034 in the left panel. To rectify the effect of the number of neurons on the decoding performance of
1035 the classifier, decoding accuracy of the classifier trained with a balanced number of randomly
1036 chosen neurons was visualized as control. Data are mean with \pm SEM and gray dots represent
1037 predictions from each imaging session. Two-way ANOVA. Šídák's multiple comparison
1038 between the decoding accuracy trained without highly tuned neurons and the control.

1039

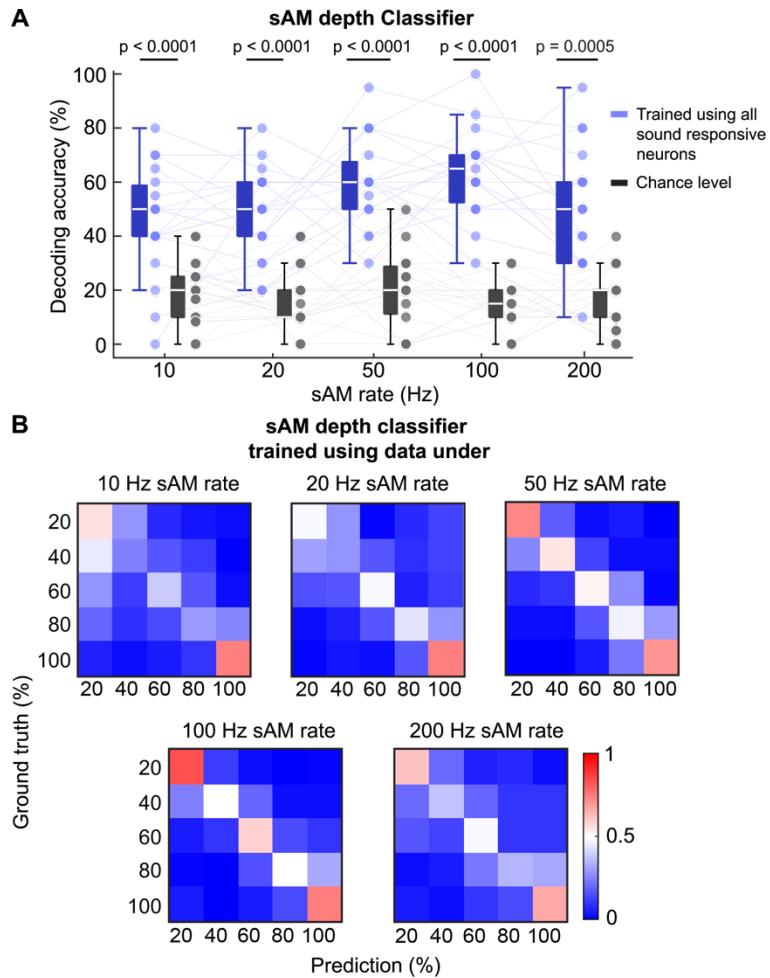


1040

1041 **Figure 5**

1042 Decoding sAM rate using sound-excited and sound-inhibited neurons. **A.** Decoding accuracy of
 1043 sAM rate classifier trained using sound-inhibited neurons and corresponding chance level. Two-
 1044 way repeated-measures ANOVA. Šídák's multiple comparison between decoding accuracy
 1045 trained using sound-inhibited and chance **B.** Decoding accuracy of sAM rate classifier trained
 1046 using sound-excited neurons and corresponding chance level. To ensure an equal number of
 1047 sound-excited and sound-inhibited neurons, we randomly selected a subset of sound-excited
 1048 neurons from each imaging session. Two-way repeated-measures ANOVA. Šídák's multiple
 1049 comparison between decoding accuracy trained using balanced sound-excited and chance. **C.**

1050 Decoding accuracy of sAM rate classifier trained using all sound-responsive, sound-excited, and
1051 sound-inhibited neurons under 100 % sAM depth. Kruskal-Wallis test. Dunn's multiple
1052 comparison between decoding accuracies trained under different conditions.

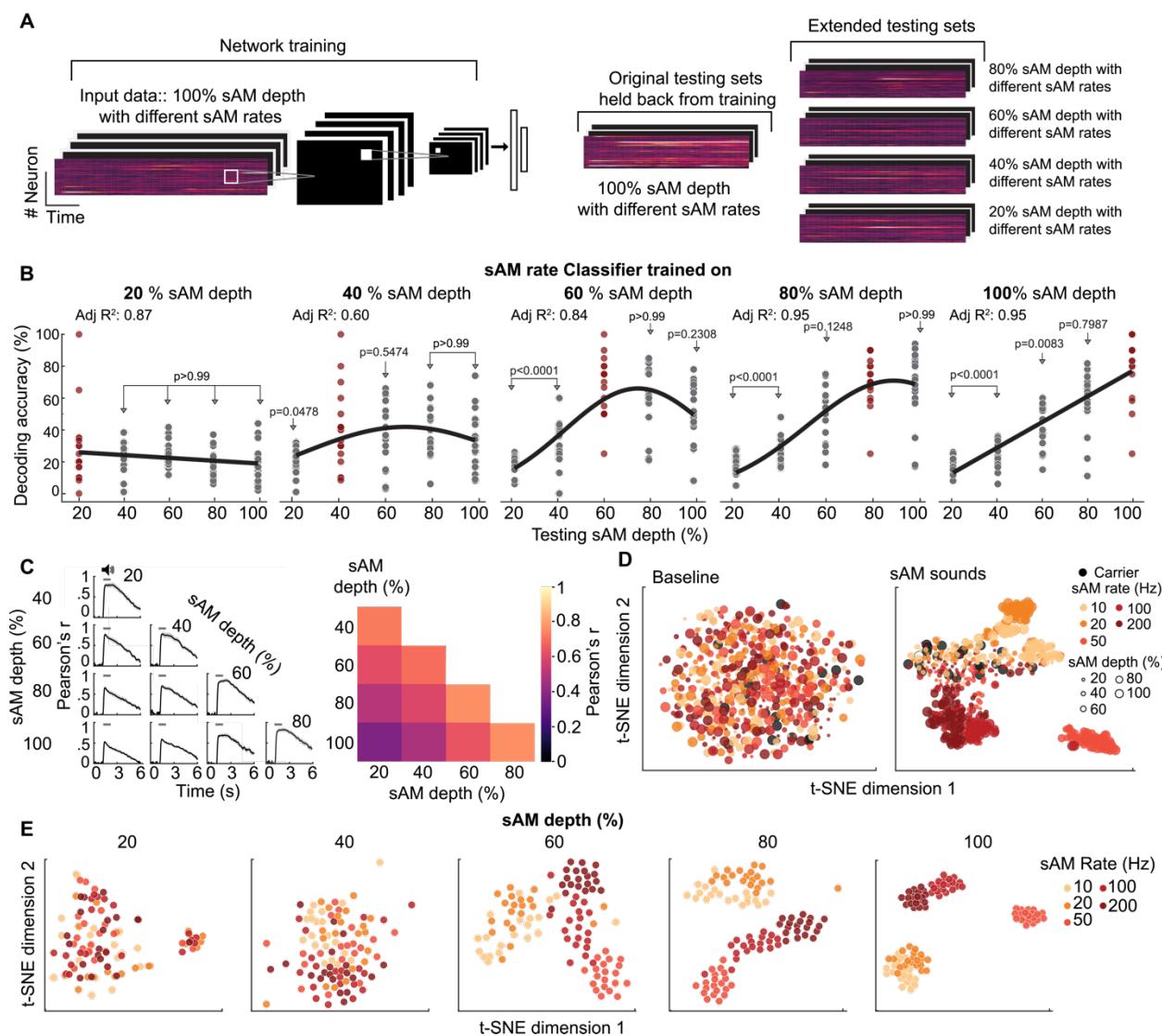


1053

1054 **Figure 6**

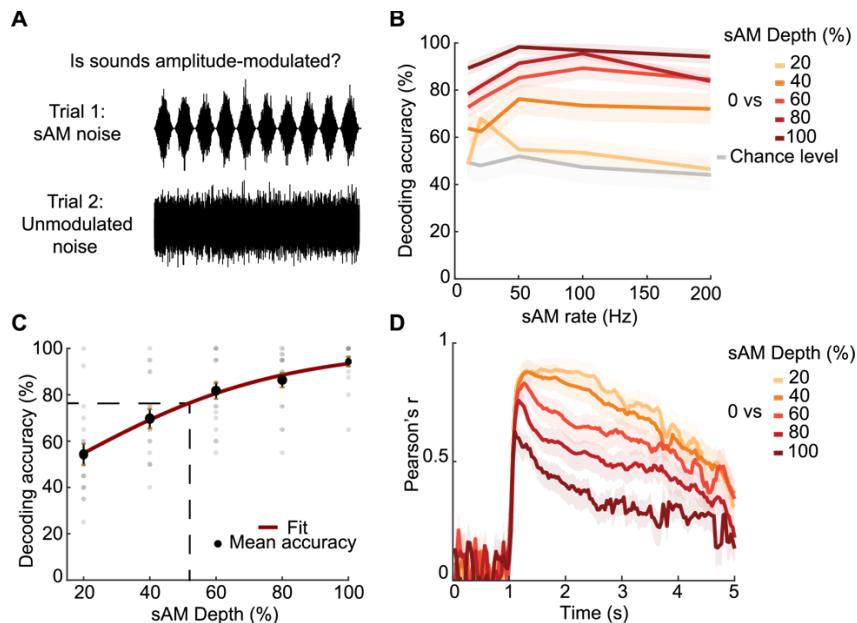
1055 Decoding sAM depth. **A.** Decoding accuracy of the sAM depth classifier under a given rate and
1056 its corresponding chance level. Two-way repeated-measures ANOVA. Šídák's multiple
1057 comparison between decoding accuracy trained using all neurons and the chance. **B.** Normalized
1058 confusion matrix of sAM depth classification under a given rate averaged across imaging
1059 sessions.

1060



1063 Representation of different sAM depths in the shell IC. **A.** For sAM rate classification, the
 1064 decoder was trained using input data from a specific sAM depth. After training was complete,
 1065 the decoder was evaluated on both original testing sets held back from training and extended
 1066 testing sets (all datasets from other sAM depths). **B.** Evaluation of the sAM rate classifier on
 1067 extended testing sets (black) and held-back testing sets (red). The curve was fitted using a
 1068 gaussian model. Friedman test for each sub-panel. Dunn's multiple comparison between
 1069 decoding accuracies tested on held-back and extended testing sets. **C.** Pattern correlation of each

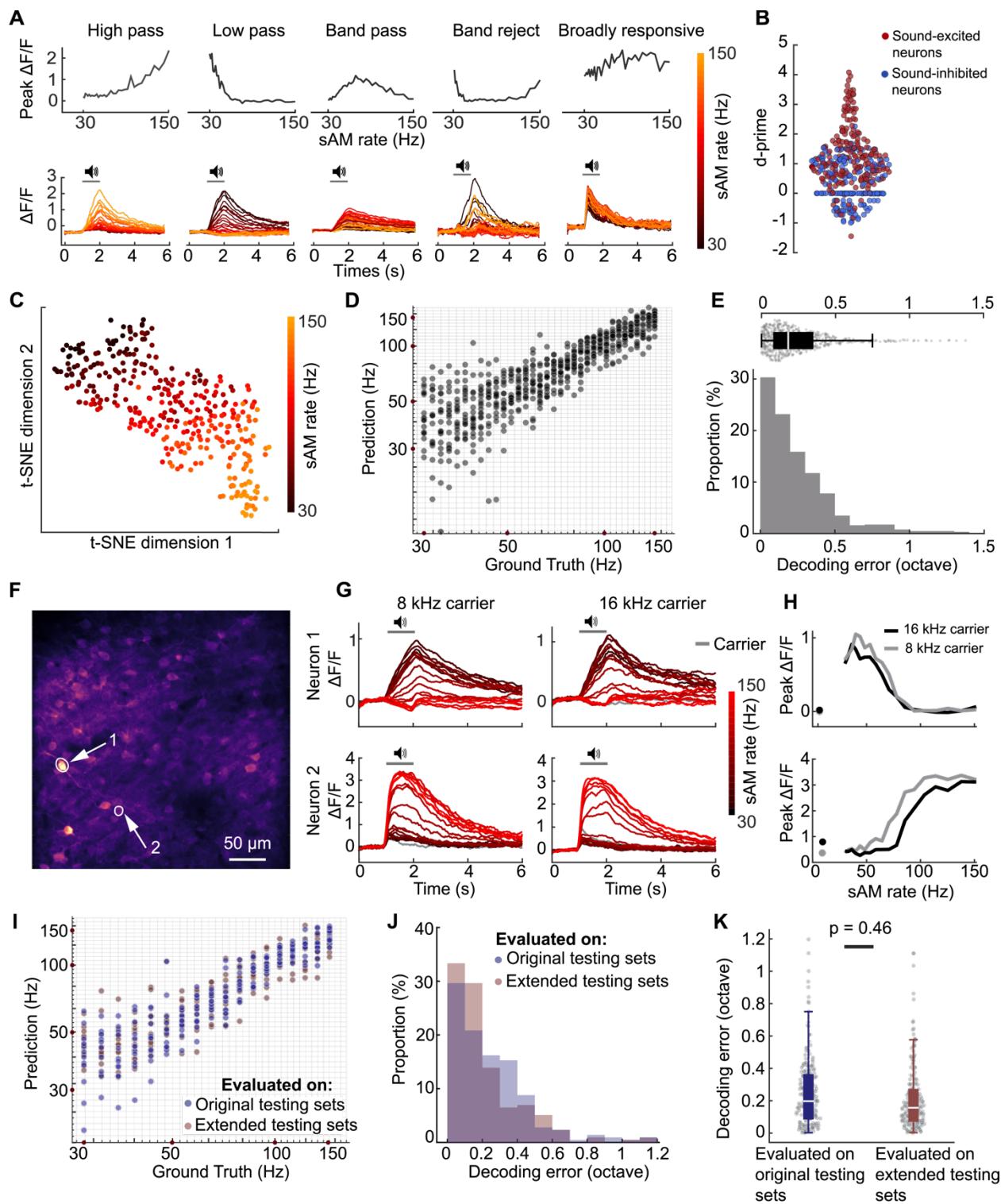
1070 pair of sAM depths. Left, Correlation of the trial-averaged neural population vector across two
1071 different sAM depths or -rates on a per-frame basis. Data are mean \pm SEM Right, averaged
1072 correlation data during sound presentation. **D.** Two-dimensional t-SNE map of shell IC neural
1073 population activity. Left, t-SNE map of neural population activity during baseline. Right, t-SNE
1074 map of neural population activity during sound presentation. **E.** t-SNE map of neural population
1075 data with a single sAM depth only. Each dot represents a single trial neural population activity.



1076

1077 **Figure 8**

1078 Binary classification of sAM and unmodulated sounds. **A.** Binary classification paradigm: A
1079 CNN decoder was trained to classify sAM (with different sAM depths and -rates) and
1080 unmodulated noise. **B.** Decoding accuracy of binary classification of sAM sounds and
1081 unmodulated noise and corresponding chance level. **C.** Logistic curve fitting of the binary
1082 classification performance. Dashed lines, both horizontal and vertical, denotes the position of the
1083 half maximum on the fitted curve. **D.** Pattern correlation between sAM sounds with different
1084 sAM depths and unmodulated sounds. For **B** and **D**, data are mean \pm SEM.



1085
1086

Figure 9

1087 Neuronal responses to narrowly spaced sAM rates in the shell IC. **A**. Examples of modulation
1088 transfer functions to fully modulated sAM sounds with 30-150 Hz sAM rates. Top: peak

1089 response of example neurons at varying sAM rates and 100% sAM depth. Bottom: trial-averaged
1090 fluorescence traces of example neurons. **B.** Distribution of d-prime of shell IC neurons. **C.** t-SNE
1091 map of shell IC population responses to narrowly spaced sAM rates. Each dot represents neural
1092 population activity in a single trial. **D.** Regression performance of CNN decoder using
1093 population fluorescence data of shell IC neurons with narrowly spaced sAM rates and 100%
1094 sAM depth. Each dot is the prediction of a single trial in testing sets. The coordinates are plotted
1095 on a logarithmic scale of base 10. **E.** Distribution of decoding errors in octaves. **F.** Example
1096 imaging FOV. **G-H.** trial-averaged responses (**G**) and peak of response (**H**) of two example
1097 neurons to sAM sounds with 16kHz and 8kHz center frequencies for the noise carrier. **I.** sAM
1098 rate decoding performance: sAM rate decoder was trained using data from sAM sounds with
1099 central frequency of either 8 kHz or 16 kHz for the noise carrier. After training, the decoder was
1100 evaluated on both original held-back testing sets with the same center frequency as in the
1101 training set, and extended testing sets with a different central frequency of the carrier. **J-K.**
1102 Distribution of decoding error in octave. Wilcoxon signed-rank test.

1103 **Table 1.**

1104 Behavioral threshold for sAM rate discrimination of fully modulated sAM sounds across
1105 different species.

1106

Source	Subject	sAM rate range	Threshold (Δ sAM rate)	Performance	Task
Dooling and Searcy (1981)	Parakeet	40-160 Hz	5-12 Hz	—	Detection of increment of sAM rate
Long and Clark (1984)	Chinchilla	20-160 Hz	8-40 Hz	—	Detection of increment of sAM rate
Long and Clark (1984)	Human	20-160 Hz	2-10 Hz	—	Detection of increment of sAM rate
Formby (1985)	Human	20-160 Hz	3-11 Hz	—	Detection of increment of sAM rate
Moody (1994)	Macaque	20-160 Hz	4-10 Hz	—	Detection of increment of sAM rate
Kelly et al. (2006)	Rat	10-100 Hz	10-45 Hz	—	Detection of increment of sAM rate
Kurt and Ehret (2010)	Mice	20 and 40 Hz	—	d-prime = 1-1.5	Two-alternative forced-choice sAM rate discrimination
Mackey et al., (2022)	Macaque	20-120 Hz	5-10 Hz	—	Detection of increment of sAM rate. Δ sAM rate was varied over a range of 1-64 Hz
Yao et al., (2023)	Gerbil	4 and 10 Hz	—	90% correct	Two-alternative forced-choice sAM rate discrimination

1107 Related to Figure 9.