Title: Spatially clustered neurons encode vocalization categories in the bat midbrain

Authors: Jennifer Lawlor[1,2], Melville J. Wohlgemuth[3], Cynthia F. Moss[1,2,4,5], Kishore V. Kuchibhotla*[1,2,4,6,7]

Affiliations
[1] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, 21218, USA
[2] Johns Hopkins Kavli Neuroscience Discovery Institute, Baltimore, MD, 21218,MD
[3] Department of Neuroscience, University of Arizona, Tucson, AZ, 85721, USA
[4] The Solomon Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA
[5] Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA
[6] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218, USA
[7] Lead contact
* Correspondence: kkuchib1@jhu.edu

Abstract

Rapid categorization of vocalizations enables adaptive behavior across species. While categorical perception is thought to arise in the neocortex, humans and other animals could benefit from functional organization of ethologically-relevant sounds at earlier stages in the auditory hierarchy. Here, we developed two-photon calcium imaging in the awake echolocating bat (*Eptesicus fuscus)* to study encoding of sound meaning in the Inferior Colliculus, which is as few as two synapses from the inner ear. Echolocating bats produce and interpret frequency sweep-based vocalizations for social communication and navigation. Auditory playback experiments demonstrated that individual neurons responded selectively to social or navigation calls, enabling robust population-level decoding across categories. Strikingly, category-selective neurons formed spatial clusters, independent of tonotopy within the IC. These findings support a revised view of categorical processing in which specified channels for ethologically-relevant sounds are spatially segregated early in the auditory hierarchy, enabling rapid subcortical organization of call meaning.

Introduction

Vocal communication exemplifies the ecological importance of categorical perception. Humans readily recognize the identity of a single person despite similarity in spectrotemporal features among many individuals. Moreover, categorical discrimination of vocal communication calls appears to be phylogenetically conserved across species, including those that communicate on land (non-human primates[1], naked mole rats[2], mice[3], crickets[4], frogs[5]), sea (dolphins[6], pinnipeds[7]), and air (songbirds[8], echolocating bats[9]). Current models of categorical perception propose that sensory cortical networks are the engine underlying category formation[10] – binding relevant features across a basis set of lower-level features that are initially decomposed from natural scenes by sensory organs and processed hierarchically in a feedforward manner [11–13]. While the perception of categories may arise in the cortex, categorical representations at earlier stages in sensory processing would provide significant computational advantages, conferring increased speed and fidelity of categorization.

To what extent do categorical representations emerge earlier in the sensory hierarchy? Categorical representations may exist in the form of single neuron selectivity, population selectivity, or even in spatially segregated ensembles[14]. We reasoned that applying an optical approach, which provides high spatiotemporal resolution, in a mammalian species that exhibits a rich repertoire of vocalization categories would allow us to test these possibilities. The big brown bat, *Eptesicus fuscus*, is an insectivorous bat species that stands out in the animal kingdom for its acoustic repertoire for echolocation-based navigation[15] and for social interactions with conspecifics[16,17]. *E. fuscus* moves in three dimensions through its environment and must rapidly distinguish between vocalizations produced for navigation or social interaction to successfully hunt and communicate.

Results

Two-photon calcium imaging in the awake, echolocating bat

One candidate subcortical region is the inferior colliculus (IC) because of its key position as an obligatory auditory processing station in the central nervous system. It has multiple subdivisions that receive inputs from the auditory brainstem [18]. One subdivision, the dorsal cortex of the IC (DCIC), is of particular interest for categorical representations as it receives both feedforward input and significant top-down input from the auditory cortex[19,20]. Here, we developed
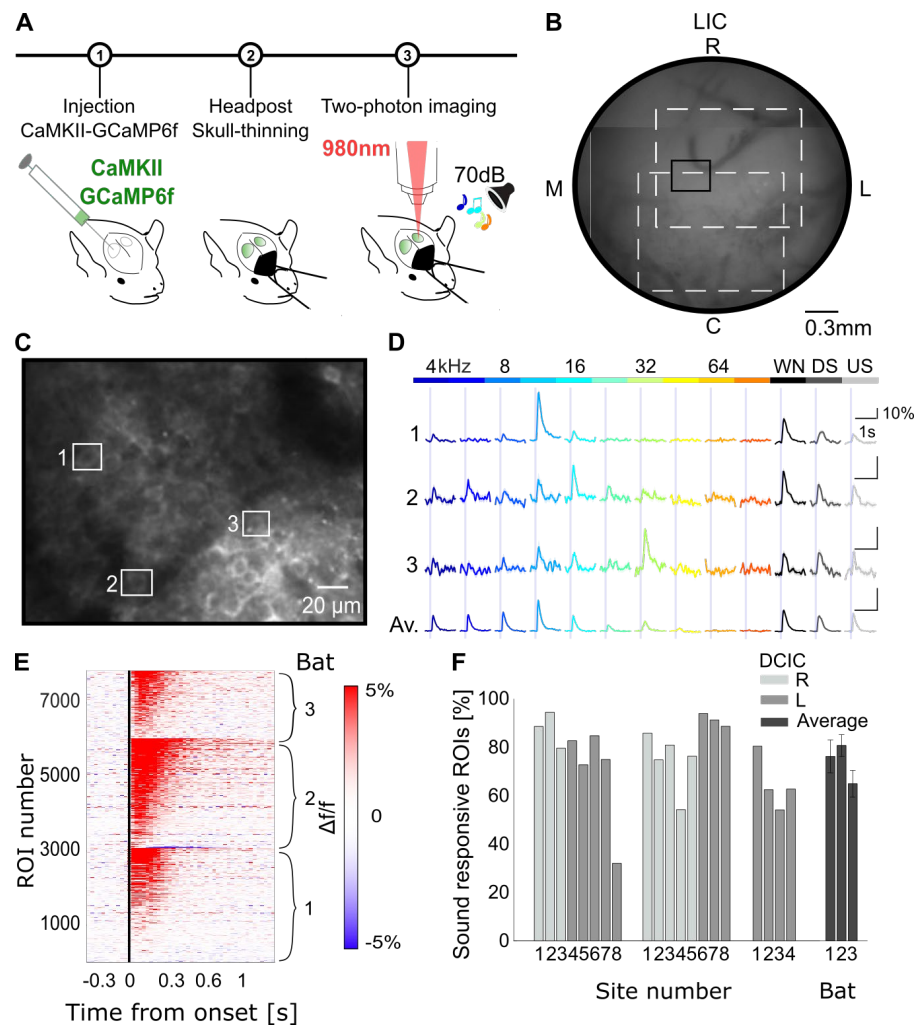
Figure 1. Robust sound-evoked responses in the DCIC of awake bats observed using two-photon calcium imaging. A. In vivo two-photon calcium imaging methods and timeline graphical representation. **B.** Composite vasculature map of an example LIC with overlaid imaging sites at 2X magnification (white box). The small black box highlights the position of the cells displayed in C. **C** Example site mean image cropped. White boxes surround examples cells presented in D. **D.** Average activity (Δf/f) of 3 example ROIs to pure tones and complex sounds presentations (color indicates stimulus identity displayed above, shading: SEM, WN: white noise, DS: frequency modulated down sweep, US: frequency modulated up sweep). The fourth calcium trace is the average trace for the imaged population of the example site (Av., shading: SEM). **E.** Sound-evoked heatmap for all identified ROIs across all sites and bats (n_{cells} = 7868, averaged over all stimuli types presented in this manuscript). **F.** Percentage of significantly sound-evoked ROIs per site (light grey: right DCIC, medium grey: left DCIC) and averaged across site (darker grey) for each bat.

65   an approach for two-photon imaging in the inferior colliculus (IC) of the awake, big brown bat, *E.*

66   *fuscus* (**Figure 1**) to study spatially resolved single neuron and population coding of auditory

67   stimuli**.** We first sought to characterize the sound-evoked properties of DCIC neurons and validate

68   the use of two-photon calcium imaging in this non-traditional animal model. We injected AAV5-

69   CAMKII-GCaMP6f bilaterally into the DCIC of *E. fuscus* and used a thinned-skull procedure to

70   gain chronic optical access (**Figure 1A-B**). We used two-photon imaging to monitor large

71   populations of excitatory neurons (**Figure S1A-B**) with single-cell resolution (**Figure 1C**). We

72 imaged from multiple sites in the left and right IC (LIC and RIC, **Figure 1B**), and at multiple

73 depths, from the brain's surface and up to 200 μm. Using this technique, we recorded neural

74 activity from 7,868 spatially-resolved neurons across 20 sites and 3 bats. Awake bats were

75 presented with an array of auditory stimuli including pure tones (4-100 kHz), complex stimuli (FM

76 sweeps and white noise), and a series of bat and mouse vocalizations. Individual neurons exhibited

77 robust sound-evoked responses (**Figure 1C-D**) with most of the recorded population responding

78 to at least one of the stimuli presented (**Figure 1E-F**, n=3 bats, 79.6±3.7%, p < 0.05, paired t-test).

79

80 **Functional microarchitecture of the DCIC**

81 We next sought to understand the functional microarchitecture of the DCIC in echolocating

82 bats. A fundamental feature of the auditory system is the precise topographic organization of sound

83 frequency, commonly referred to as 'tonotopy'. Tonotopy emerges from the biomechanical

84 properties of the basilar membrane in the inner ear and propagates throughout the central auditory

85 system[21]. The central nucleus of the IC (CNIC) in bats and other mammalian species exhibits a

86 dorso-ventral tonotopic axis[22] but less is known about the topographical organization of sound

87 frequency in the DCIC, in part because prior studies in the bat have not consistently identified the

88 precise recording site [23,24]. We presented pure tones ranging from 4 to 90 kHz, matching *E. fuscus'*

89 hearing range[25], at 70 dB with a half-octave spacing and then calculated single-neuron tuning

90 curves to identify a neuron's 'Best Frequency' (BF), here defined as the sound frequency that

91 elicits the highest amplitude stimulus-evoked response. Most neurons were significantly tuned to

92 a limited band of sound frequencies (66.4±11%, two-way ANOVA for tone and duration, p<0.05,

93 **Figure 2A-C**) in the low-to-mid frequency range (**Figure 2C**, 4 to 32 kHz: 93.3%±1.7% of tuned

94 neurons). These findings are consistent with those reported in the mouse using two-photon calcium

95 imaging[19,26]. In addition, these data are consistent with anatomical evidence showing that antero-

96 ventral cochlear nucleus neurons that encode lower-frequency sounds project more superficially

97 than dorsal axonal projections[27], suggesting a conserved functional microarchitecture across

98 species.

99 We then used the spatial resolution of two photon imaging to reconstruct the precise

100 location of each neuron in three dimensions with its corresponding BF (**Figure 2D**). We aligned

101 across imaging sites and bats in the rostrocaudal and the mediolateral extents, based on a common

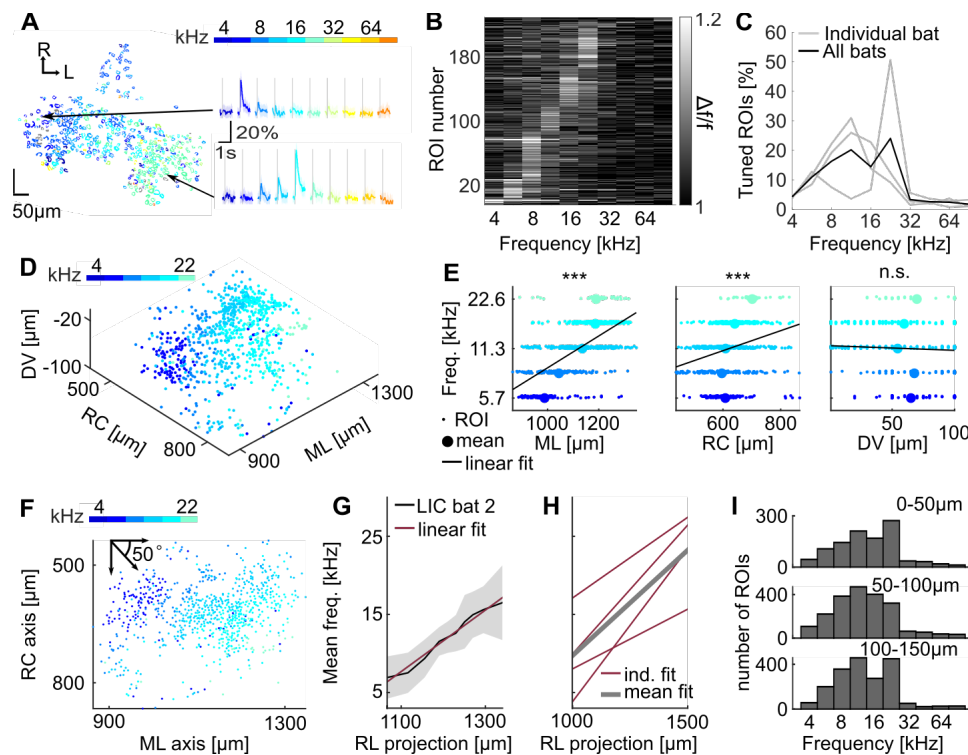102 anatomical landmark, the intersection of the IC, superior colliculus and midline. Z-depth was

Figure 2. Superficial tonotopy in the DCIC. A. Left: example site from the RIC of bat 2, ROIs color-coded by their BF. Right: tone-evoked PSTHs for example cells. The first cell showed clear tuning to 5.7 kHz and the second to 16 kHz. **B.** Tuning curves for all tone-responsive ROIs from the example site presented in A. **C.** Distribution of tuning per bat (grey, n = 3) and averaged (black) as a percentage of the tuned population. Most of the ROIs' tuning is in the low to mid-frequency range (93.3±1.7% of tuned neurons' BF between 5.7 to 22.6kHz) **D**. 3D reconstruction of bat 2 LIC (n_{site} = 3, n_{cells}= 835), tuned ROIs are color-coded according to their BF, showing low frequency rostrally and medially and higher frequency caudally and laterally **E**. ROIs' BF (individual ROI: small dot, average: larger dot) as a function of distance along three along three anatomically defined axis for example LIC in D for each frequency (left: ML, middle: RC and right: DV). Linear fit (black line) is significant for ML and RC but not for DV (p=5.02e-95, p=4.79e-15 and p =0.13, respectively) **F.** 2D reconstruction of bat 2 LIC, tuned ROIs are color-coded according to their BF showing frequency bands oriented rostromedially. **G.** Mean BF (in black) and its linear fit (in red) along the 50⁰ RL axis for LIC in F. Mean tuning increases along this axis, confirming the existence of a tonotopic gradient. **H.** Linear fit (in red) of tuning along the 50⁰ RL axis for each bat hemisphere (in red) and the average (in grey). Tuning increases as a function of RL distance by 2.1oct./mm on average. **I.** Frequency tuning distribution across bats along the DV axis (50 µm bins from 0 to 150 µm). Tuning does not significantly vary with depth (Friedman test, p_{dv}=0.4066).

103 measured using the optical sectioning capabilities of two-photon imaging coupled with an estimate

104 of the skull-to-brain transition, based on a change in background fluorescence (**Figure S2A-C**).

105 This allowed us to test whether a spatial, tonotopic gradient exists in the mediolateral, rostrocaudal,

106 or dorsoventral axes. We determined that a positive linear relationship exists for both the

107 mediolateral and the rostrocaudal axes but not for the dorsoventral axis (**Figure 2E**, p_{ML} = 5.02e-

108 95, p_{RC}= 4.79e-15, and p_{DV} = 0.13, additional example sites in **Figure S3A**). To capture the direction

109 of this tonotopic axis, we estimated the best fit angle per reconstructed hemisphere across bats

110 (**Table S1**), yielding an average angle of 47.5⁰, a value remarkably close to the 50⁰ reported in the

111 mouse DCIC [26]. The tonotopic gradient was computed as the linear fit of the mean frequency along

112 this rostrolateral axis (**Figure 2F-G**) per hemisphere and per bat. On average, the BF increased by

5

113   2.1 oct./mm along this axis (**Figure 2G-H**). To confirm that there was no observable tonotopic

114   organization in the dorsoventral axis, we compared tuning at three different depths and observed

115   similar frequency tuning distributions (**Figure 2I**, Friedman test, p_{dv}=0.4066).

116

Neural selectivity, population decoding, and spatial clustering of ethologically-relevant
vocalizations in the DCIC

119   We next sought to characterize the response properties of DCIC neurons during playback

120   of natural vocalizations from conspecifics. Bats signal to other bats with a frequency-modulated

121   bout [9,28] to claim food when competing with conspecifics ('social' vocalizations), while also using

122   echolocation to track prey and steer around obstacles ('navigation' vocalizations)[29,30]. These

123   vocalizations exhibit largely overlapping spectrotemporal features with subtle differences, such as

124   shallower FM sweep rate and greater low frequency energy of social calls compared to navigation

125   calls [30] (**Figure 3A**). The vast majority of recorded neurons responded to conspecific vocalizations

126   (86±18.4%, ANOVA for responsivity, p<0.05, **Figure S4A**) whose spectral content spans from 22

127   to 90 kHz, highlighting a discrepancy between responsivity to vocalizations versus pure tones: in

128   this population, neurons' BF's were generally below 32 kHz (**Figure 2C**). The population sound-

129   evoked response was highest for conspecific vocalizations and lowest for pure tones (**Figure 3B**).

130   These data suggest that ethologically relevant vocalizations drive DCIC neurons more than pure

131   tones or even heterospecific complex stimuli, such as, mouse ultrasonic vocalizations ('mouse

132   USV', **Figure 3A**). This result is consistent with electrophysiological recordings in a number of

133   species showing that conspecific vocalizations elicit larger evoked responses than more artificial

134   stimuli in the IC[31]. We observed a striking selectivity within individual neurons for categories of

135   bat vocalizations (social: 30.2±4.5%; navigation: 19.7±3.7%) but little to no selectivity for mouse

136   USVs (**Figure 3D,** p=0.13, KW test vs. pure tones**,** detailed selectivity in **Figure S4B**),

137   highlighting the importance of ethological relevance in auditory response selectivity. Given the

138   lack of selectivity to mouse USVs, we focused on the bat vocalizations for further analysis. We

139   calculated the neural selectivity to social versus navigation calls by computing the difference in

140   evoked responses between both categories (**Figure 3C-E, and S4C-D**). While there were neurons

141   selective to both bat vocalization categories (social or navigation), the DCIC network exhibited a

142   strong preference for social calls (**Figure 3E and S4D**, overall, n=1,383 social-selective neurons,

143   n=502 navigation-selective neurons; per bat, 24±5.4% social-selective and 8.4±1.9% navigation-

144   selective, KW test, p=5.3e-05)[30]. These data demonstrate that individual DCIC neurons are

145   selective to conspecific vocalizations with a strong preference for encoding social vocalizations.
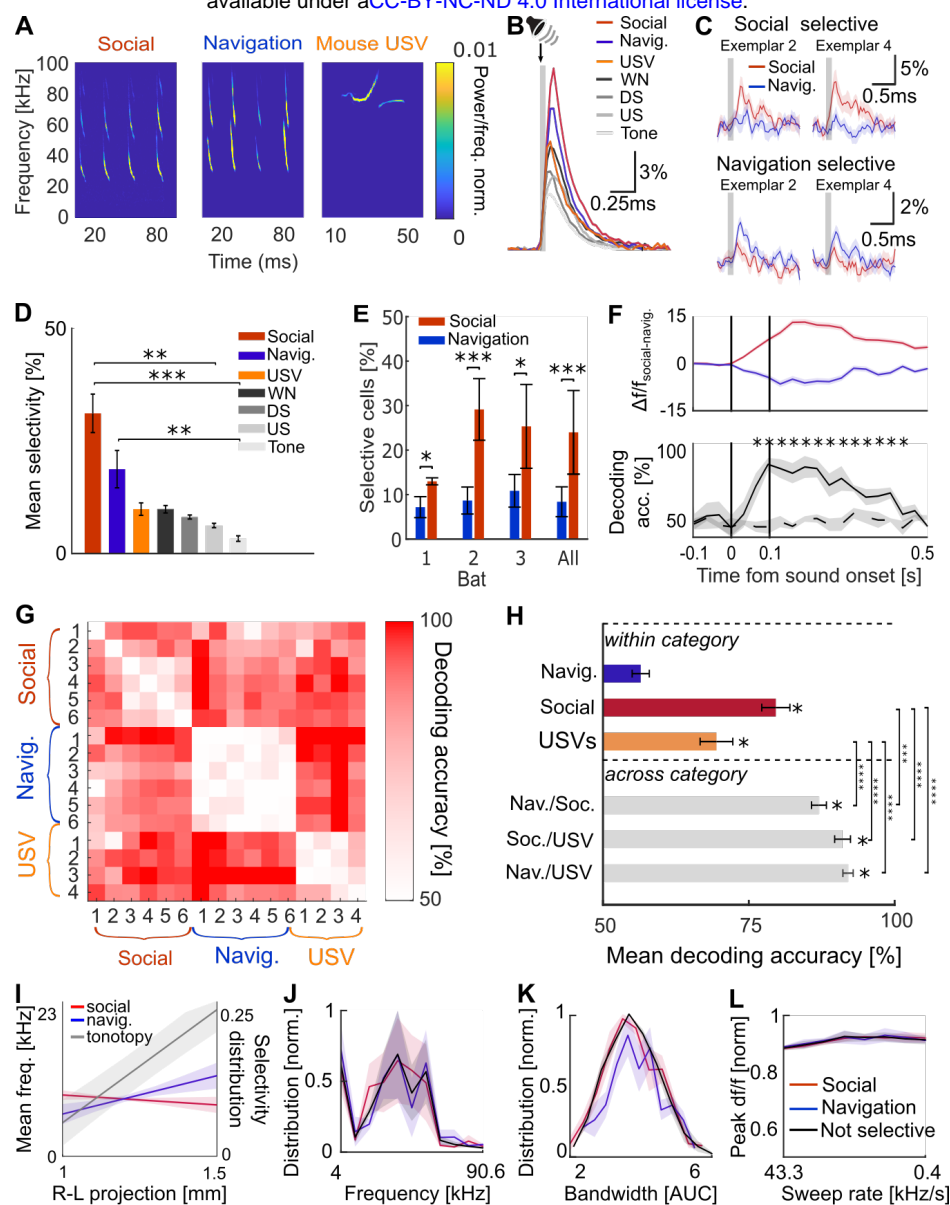
Figure 3. Categorical encoding of bat vocalizations in the DCIC. A. Example spectrograms for the vocalization categories presented. Left: social call sequence, center: temporally matched navigation call sequence, right: mouse ultrasonic vocalization (USV). **B.** Population sound-evoked response as $\Delta f/f$ for all stimulus types presented (n_{bat}=2, n_{cell}=2232). Conspecific bat calls evoked the highest evoked-response followed by mouse USVs and white noise, with tones evoking the smallest. All stimuli evoked significantly different responses (Friedman test on peak amplitude, α<0.05) apart from USVs and navigation calls**. C.** Example cells' sound-evoked responses ($\Delta f/f$, median, shading: SEM) for two exemplars of social calls (red) and corresponding navigation calls (blue). Top: example social selective cell. Bottom: example navigation selective cell. **D.** Mean population selectivity for stimulus categories. The imaged population is more selective to the bat calls than other stimulus types (KW test, $p_{social/tone}$=0.0001, $p_{navig.l/tone}$=0.0049, $p_{social/us}$=0.0146). **E.** Average percentage of selective cells per bat and across bats (error bar: STD). Neuronal population shows a bias towards social selectivity (KW test, p_{bat1} = 0.0495, p_{bat2} = 0.0008, p_{bat3} = 0.0495, p_{all} = 5.2683e-05). **F.** Time course of population selectivity. Call sequence onset and offset is indicated by the black lines. *Top*: difference between mean social-evoked and navigation $\Delta f/f$ for top 5% social selective cells (in red, average across bats, shading: SEM) and top 5% navigation selective cells (in blue). Note that the difference in activity between social and navigation preferring cells increases rapidly after stimulus onset. *Bottom*: Time course of decoding accuracy (social vs navigation, mean across bats = plain line, shading = SEM), star indicates significance compared to chance level decoding (dashed line, permutation test, α<0.05). Bat call category can be decoded from neuronal population 60ms after sound onset before the end of the call sequence. **G.** Pairwise decoding accuracy averaged across bats for each call (conspecific or heterospecific) exemplar pairs. Decoding accuracy within categories is lower than across category (white = chance, red = 100% accuracy). **H.** Mean decoding accuracy within and across categories (error bar: SEM). Higher than chance decoding accuracy is indicated by a single star above each bar. Decoding accuracy is significantly lower vs. across categories (ANOVA, $p_{S/N/S}$=0.0001, $p_{S/S/U}$<e^{-4}, $p_{U/N/S}$<e^{-4}, and $p_{U/S/U}$<e^{-4}). Note that decoding accuracy is not significantly different between groups within (ANOVA, $p_{S/U}$= 0.0808) and groups across (ANOVA, $p_{N/S/S/U}$ = 0.4619, $p_{N/S/N/U}$ = 0.2436, $p_{S/U/N/U}$ = 0.9999). **I.** Linear fit of cell ratio along the tonotopic gradient (gray) for both social selective (red) and navigation selective (blue). Averaged cell ratio for both social and navigation selective cells are significantly different than the tonotopic gradient (ANOVA, p <e^{-4}) but not from one another (ANOVA, p = 0.9979). **J.** Mean best frequency normalized distribution for social, navigation and non-selective cells (shading: SEM, Friedman, p = 0.9688). **K.** Mean bandwidth normalized distribution for social, navigation and non-selective cells (shading: SEM, ANOVA, p = 0.8182). **L.** Mean downsweep rate tuning as peak $\Delta f/f$ normalized for social, navigation and non-selective cells (shading: SEM, ANOVA, p = 0.9425).

146

147   Bat vocalizations tend to be grouped in sequences, for example the social calls presented

148   here contain 3 to 4 successive calls. Therefore, categorical information could be carried not only

149   by the spectral content but also by the temporal pattern of call sequences. To determine if the

150   selectivity observed is due to the processing of the spectral information of the first call, the

151   processing of the entire sequence or a mixture of both, we analyzed the time course of the neural

152   selectivity. The activity of selective neurons diverged rapidly after the sound onset (**Figure 3F**,

153   top panel) and before the end of the call sequence. A linear decoder reliably discriminated between

154   conspecific categories on the second imaging frame after sound onset (~32 ms, 31.25 Hz

155   acquisition rate, **Figure 3F**, bottom, **Figure S5B** for comparison with SI), arguing that the

156   selectivity and population-level representations arise within one to two calls.

157   The high proportion of social-selective neurons and rapid population discriminability

158   suggests that the DCIC network may encode social vocalizations with high fidelity. To test this,

159   we shifted from single-neuron analyses to population decoding. We trained a linear decoder for all

160   pair-wise stimuli presented to the bats and found that decoding within category performed best for

161   social vocalizations (**Figure 3G-H, Figure S5C** for comparison with SI), suggesting that the

162   higher proportion of social-selective neurons allows the network to encode these vocalizations

163   with higher fidelity (even if the spectral distance within social calls is lower than within mouse

164   USVs, **Figure S5A**). We then sought to test the extent to which the DCIC network encoded call

165   category by looking at decoding strength within versus across categories. Interestingly, decoding

166   accuracy was significantly higher across than within categories for all category types (**Figure 3G-**

167   **H**, ANOVA, $p<10e^{-5}$ and p = 0.0001). Despite the lack of single neuron selectivity to mouse

168   USV's, the DCIC network reliably decoded individual calls within the mouse USV category

169   (70±2.5%, **Figure 3G**), but to a lesser extent than across categories of bat vocalizations (91±1.4%,

170   **Figure 3G**) which was similar to conspecific categorical decoding (**Figure 3F**, 87±1.4%,

171   91.1±1.4%, and 91.9±0.9% respectively, ANOVA, $p_{con/social/USVs}$ = 0.4619, $p_{con/echo/USVs}$ = 0.2436).

172   The similarity in cross-category decoding between conspecific vocalizations (social vs.

173   navigation) and conspecific vs. heterospecific (social vs. mouse UV, navigation vs. mouse USV),

174   is at odds with the sounds' spectral distances (**Figure S5A**) and indicates that the DCIC network

175   contains a higher-fidelity representation of conspecific calls, likely relying on species-specific

176   spectrotemporal filtering properties.

177   We next sought to understand whether neurons' selectivity to distinct conspecific
178   vocalizations could be explained by differences in lower-level auditory feature responses. For
179   example, social calls have shallower frequency modulated component at the lower frequency tail
180   of the sweep compared to navigation calls, and social-selective neurons may more strongly encode
181   low frequency stimuli, clustering along the tonotopic axis to reflect their tuning preference.
182   Surprisingly, this was not the case. The ratio of social-selective and navigation-selective neurons
183   was uniformly distributed along the recorded tonotopic axis (**Figure 3I**). Importantly, other low-
184   level features, including frequency tuning, bandwidth, duration and complex sound responsivity
185   were identical for social and navigation selective neurons (**Figure 3J-L**, tuning: Friedman test, p
186    $= 0.9688$; bandwidth, ANOVA, $p = 0.8182$; down sweep rate tuning, ANOVA, $p = 0.9425$; **Figure
187   S6A-C**). Thus, low-level features could not explain categorical selectivity, indicating that neural
188   selectivity for vocalizations in the DCIC may reflect a higher-order categorical representation.

189   Finally, we sought to evaluate whether neural encoding of vocalizations in the IC are
190   spatially organized by exploiting the high spatial resolution of two-photon imaging. For each
191   category-selective neuron, we computed the percentage of neighboring neurons that were also
192   selective to the same category in 30 µm steps in a site-specific manner. Surprisingly, this revealed
193   clusters of category-selective neurons to either social or navigation calls for all bats in both
194   hemispheres (**Figure 4**). The percentage of social selective neighboring cells is significantly higher
195   than what would be expected by chance from a 'spatial shuffle' for distances of 30 µm and 60 µm
196   (**Figure 4B-C**, $p_{30} = 0.01$, $p_{60} = 0.01$, see inset, ANOVA, $p_{30-90} = 1.2692e-07$). The same
197   relationship was observed in navigation selective cells (**Figure 4D-F**, $p_{30}=0.01$, see inset,
198   ANOVA, $p_{30-90} = 1.2773e-04$). To further describe the spatial organization of these selectivity
199   clusters, we identified cluster centers and used a composite map to compare their location with the
200   anatomically defined tonotopic gradient (**Figure 4D, H**). Clusters centers are distributed along the
201   tonotopic axis and follow the cell density distribution (**Figure S7A-B** for example site and **Figure
202   S7C** for cluster distribution). Neurons exhibiting call selectivity were, thus, spatially clustered into
203   functional microdomains largely independent of the larger-scale tonotopic gradient.

204

205   **Discussion**

206   Current models of categorical perception rely on the idea that the periphery and midbrain
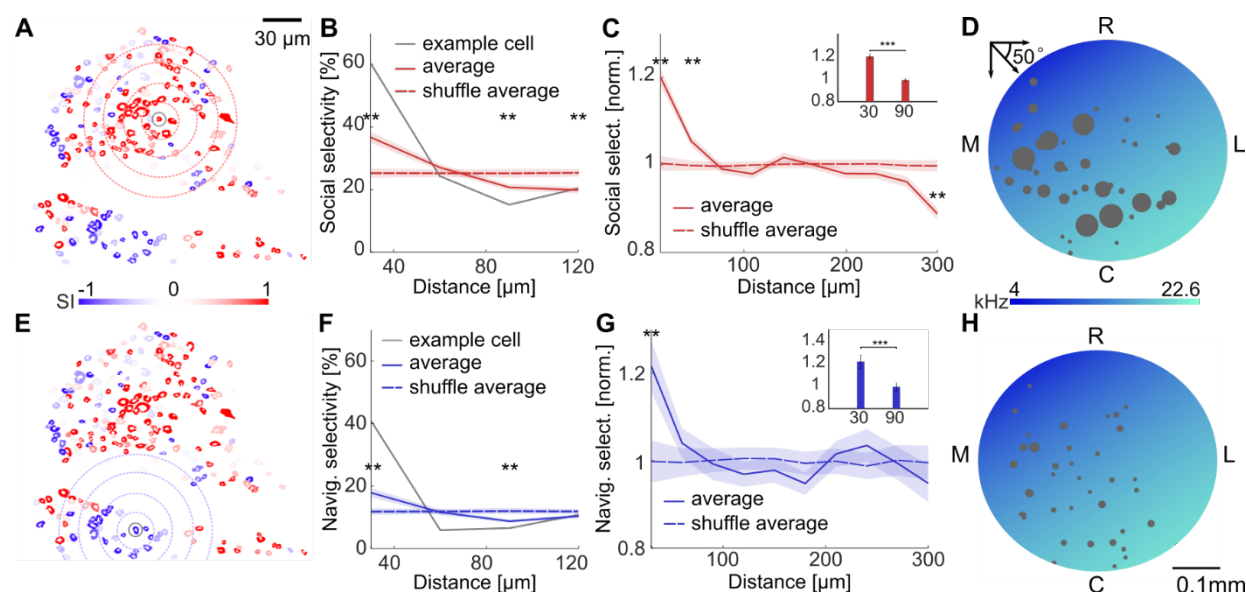207   serve primarily a feedforward and filter-bank role[10,32,33]. However, our data demonstrate that

10

Figure 4. Call selectivity is spatially clustered, independent of tonotopic gradient A. Example site, ROIs color-code corresponds to the social (red) vs navigation (blue) selectivity index. The social selective example cell presented in B is highlighted with a gray circle. Red dotted circles indicate areas used to compute the percentage of social selective neighboring cells (in 30 μm steps). **B**. Percentage social selective cells as a function of distance from social selective cells (shading: SEM) for example site in A. Example cell highlighted in A (gray line), site average (red plain line) is significantly higher than the spatial shuffle (red dotted line) at 30 μm (p = 0.01) and decreases below the shuffle line above 90 μm (p_{90} = 0.01, p_{120} =0.01). **C**. Percentage of social selective cells as a function of distance from social selective cells normalized across all sites and bats. Average is higher (red plain line) than spatial shuffle (red dotted line) at distances below 60 μm (p_{30} = 0.01, p_{60} =0.01) (shading: SEM) and lower at 300 μm (p_{300} = 0.01). Additionally, percentage (normalized) decreases significantly between 30 and 90 μm (ANOVA, p = 1.2692e-07), as displayed in the inset. **D**. Social cluster centers (gray dots, dot size is proportional to number of cells in the cluster) are distributed across composite tonotopic gradient (BF color-coded). **E**. Example site in A. The gray circle highlights navigation selective cell presented in F. Here, concentric blue dotted circles indicate the distance in 30 μm steps to the example cell. **F**. Percentage of navigation selective cells as a function of distance from navigation selective cells (shading: SEM) for example site in D. Example cell highlighted in D (gray line), site average (blue plain line) is significantly higher than the spatial shuffle (blue dotted line) at 30 μm (p = 0.01) and decreases below the shuffle line above 90 μm (p_{90} = 0.01). **G**. Percentage of neighboring navigation selective cells as a function of distance from navigation selective cells normalized across all sites and bats. Average is higher (blue plain line) than spatial shuffle (blue dotted line) below 30 μm (p_{30} = 0.01) (shading: SEM). Additionally, percentage (normalized) decreases significantly between 30 and 90 μm (ANOVA, p = 1.2773e-04), as displayed in the inset. **H.** Navigation cluster centers (gray dots, dot size is proportional to number of cells) are distributed across composite tonotopic gradient (BF color-coded).

208   categorical representations of call meaning at the single-neuron and population level emerge

209   earlier in the auditory hierarchy than posited by traditional models and display 'hotspots' of

210   spatially clustered, categorical encoding. Future work will determine the extent to which this

211   property is locally computed by the IC, at earlier stations of the brainstem, or alternatively, relies

212   on top-down input from the neocortex. This ethologically relevant neural code for categorical

213   information is independent of the superficial tonotopy in the DCIC. Our data support a revised

214   view of categorical perception in which ethologically relevant sensory streams are instantiated at

215   the single-neuron and population-level in early auditory centers to provide specified and spatially

216   segregated acoustic channels of categorical information to recipient regions. This functional

217 microarchitecture likely serves to increase the speed of transmission (by virtue of being early in
218 the sensory pathway) and reduce the wiring costs (through spatially organized channels) for
219 echolocating bats, and these computational principles likely extend to other species that extract
220 sound meaning from rich vocal repertoires.
221

Acknowledgements

228

Author contributions

230 Conceptualization: J.L., M.W., C.M., K.K.; experiments: J.L., M.W.; data analysis: J.L., K.K.,
231 writing - original draft: J.L., K.K.; writing – review & editing: J.L., M.W., C.M., K.K.
232

Declaration of interests

234 The authors declare no competing interests.

235 **Materials and Methods**

236 **Key resource table.**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
pENN.AAV.CamKII,GCaMP6f.WPRE.SV40	Addgene	100834-AAV5
Experimental models: Organisms/strains		
E. fuscus	Wild caught	N/A
Software and algorithms		
Matlab	The MathWorks	https://www.mathworks.com
ImageJ	Schneider et al.[34]	https://imagej.nih.gov/ij/

237
238 **Animals.** Data was collected from three adult big brown bats (*Eptesicus fuscus*, one male and two
239 females, ages unknown, wild caught) collected from exclusion sites in the State of Maryland. The
240 bats were collected under the permit issued by the Maryland Department of Natural Resources (no.
241 55440). All procedures were approved by the Institutional Animal Care and Use Committees at
242 Johns Hopkins University (protocol no. BA17A107), where this research was conducted.

243
244 **Anesthesia and pain management.** All surgical procedures were performed under Isoflurane
245 anesthesia (1.5%-3%) with 0.8 oxygen flow in accordance with the protocol no. BA17A107. 30
246 minutes prior to anesthesia induction, animals received half a dose of meloxicam (Metacam) orally
247 for its analgesic effect. The second half dose was delivered once the animal awoke, as well as
248 sulfamide, for antibiotic coverage. Post-surgery recovery lasted 4 to 5 days with a daily dose of
249 meloxicam and sulfamide.

250
251 **GCaMP6f viral transfection.** We injected AAV5-CamKII-GCaMP6f (Addgene, 100834-AAV5)
252 bilaterally in the IC. We first resected the muscle located on top of the skull (*temporalis medius*)
253 working from an incision placed above the midline to uncover the IC skull surface bilaterally. To
254 avoid disrupting the recording surface, we entered the IC at an angle of 35º to 50º and traveled
255 ~1mm to reach an injection site at the center of the DCIC at a depth of -600μm (see Figure 1A for
256 visualization). We used a Hamilton syringe (10μL, 700 series, 34 gauge) preloaded with the virus
257 (1μL per hemisphere) and injected using a syringe pump (Harvard Apparatus) at 75nL/min and a
258 force of 30%. We kept the needle in place for an additional 5 minutes after injection before slowly
259 removing from the brain. Animals were allowed to recover for at least 1 month before the next
260 procedure (to allow for viral expression).

261
262 **Headpost implantation.** We resected the muscle on the rostral part of the skull. A custom built
263 flat two-pronged headpost (**Figure 1A**) was cemented with dental cement (Metabond) to the
264 cranium at a distance of at least 4mm from the IC center, identified visually. The bat was then put
265 on rest for at least one day before the thinned-skull procedure to allow for the cement to cure.

266
267 **Thinned-skull approach.** We adapted thinned-skulled approached developed in the mouse [35,36].
268 Using a surgical drill, we carefully thinned the skull above both hemispheres of the IC, without
269 piercing the skull layer. The skull was then dried, and a thin layer of transparent fast-drying

13

270 adhesive was applied above the thinned area. Bats were allowed to recover for a day prior to data
271 collection.

273 **Two-photon calcium imaging acquisition.** Two-photon calcium imaging was performed using a
274 two-photon resonant microscope (Neurolabware) equipped with a 16X objective (Nikon). Two-
275 photon fluorescence was excited at 980nm using an Insight X3 laser (SpectraPhysics). Data
276 contained in this manuscript were collected at 31.25Hz over one plane at either 4X (294 μm x 192
277 μm, cropped) or 2X (568 μm x 374 μm, cropped). Additionally, a volumetric stack (z-stack) was
278 collected at the end of each session in 2 or 5 μm steps from the skull surface up to 200 μm below
279 the surface.

281 **Two-photon calcium imaging preprocessing.** We used suite2p [37] to register and align each
282 session. We drew regions of interest (ROIs) on the session mean image using ImageJ, as automatic
283 ROI identification was perturbed by large differences in brightness related to changes in skull
284 thickness over the FOV. Fluorescent time traces of individual ROIs were extracted using a custom
285 MATLAB-based toolbox. Δf/f was computed per ROI and sound presentation by dividing the raw
286 fluorescence by the median fluorescence for a baseline period defined as the 10 frames (312.5ms)
287 that immediately preceded sound onset.

289 **Site alignment.** Each imaging site position on the surface of the IC was determined by manually
290 matching the site's vasculature with the corresponding widefield vasculature image (larger). All
291 widefield images collected for one hemisphere were then tiled using the vasculature as a landmark,
292 creating a composite map per hemisphere (example of imaging sites overlay on composite
293 widefield image in **Figure 1B**). The relative position of each imaging sites to the most anterior and
294 medial widefield image was then computed allowing us to generate ML and RC coordinates within
295 one hemisphere. Those coordinates were then placed in the bat's coordinate system by manually
296 matching the vasculature from the obtained composite maps to images covering the full extent of
297 both ICs, obtained during the skull thinning.

299 **ROIs coordinates.** Mediolateral and caudorostral coordinates for each cell were obtained by
300 aligning recording sites within hemisphere and animal (n_{bat}=3, n_{hemi} = 4, n_{sites}=20) to an anatomical
301 landmark obtained from surgery images covering both hemispheres: the rostro-medial boundary
302 between SC and IC marked by the confluence of the superior sagittal and lateral sinuses. ROI depth
303 was obtained by estimating skull-thickness above each ROI and subtracting that value from the
304 overall site depth from brain surface identified using the z-stack. Skull thickness was computed
305 for each FOV as points above a 95% fluorescent boundary based on the FOV raw fluorescence
306 distribution (skull brightness is very high, **Figure S2** illustrates this method for an example site).

308 **Auditory stimulation.** All auditory stimuli were presented contralaterally to the imaging site using
309 one free field electrostatic speaker (ES1, TDT) placed at 45⁰ angle, 5cm away and at a 0⁰ azimuth
310 from the bat's head. All auditory stimuli were presented to the awake passively listening bat using
311 a high-frequency auditory signal processor (RZ6, TDT) at a 200 kHz sampling rate. Sounds were
312 generated in TDT (pure tones, white noise, upsweep, downsweep) or were recorded as WAV files
313 and then played back via the TDT. All stimuli were calibrated at 70dB using a GRAS microphone
314 (GRAS 46DP), placed at the bat's ear putative position.

Auditory stimuli. A series of stimuli were presented: pure tones ranging from 4 to 90.51kHz, ½ octave spaced (10 tones, pseudo random presentation, durations: 2, 4, 6, 10, 20ms), white noise, upsweep and downsweep (4khz to 90.51kHz, durations: 2, 4, 6, 10, 20, 50, 100, 200ms), bat and mouse vocalizations (5 or 6 exemplars: food claiming call sequences termed FMBs [9,28,30], 5 or 6 exemplars: temporally matched echolocation call sequence[30], 2 or 4 mouse ultrasonic vocalizations[38]), and bat hunting navigation call sequences [39]. Each sound was typically repeated 10 times (1 out of 23 analyzed sessions had 5 repetitions) and the inter-sound interval was kept constant at 1.2s for all stimuli except for the longer hunting navigation sequences where it was 2.4s.

Histology At the end of the experiment, the bats were perfused transcardially with 15mL of 1% phosphate buffered saline (PBS) followed by 20mL of 4% paraformaldehyde in PBS 1%. The fixed part of the brain containing the IC was sectioned in 50 μm slices coronally. To ascertain excitatory expression of GCaMP6f viral expression, immunochemistry was performed following a double staining procedure for GFP and GAD67[40].

Data analysis. All data analysis and statistical testing were performed using custom functions written in MATLAB 2018b or 2021b (The MathWorks). Normality of data was assessed prior to statistical testing using a one-sample Kolmogorov-Smirnov test. ANOVAs were performed when the distribution was normal, while non-parametric tests were applied (Kruskalwallis or Friedman) when the data did not follow the normal distribution. The tests used to assess significance are specified in the main text and in the figure legends.

Best Frequency identification. Best frequency (BF) was defined as the sound frequency evoking the highest response measured as the median of the peak amplitude across 50 trials (10 frames from stimulus onset, 312.5ms) at the sound level presented (70dB). BF analysis was restricted to neurons that exhibited significant tone-evoked activity (paired ttest, $\alpha<0.05$).

Tonotopic gradient characterization. We first established relationship between three anatomically defined axis rostrocaudal, mediolateral and dorsoventral (RC, ML and DV, respectively) and best frequency tuning (limited to the 4-22.7kHz range that comprises most of the tuning, **Figure 2C**) per hemisphere and animal, using a linear regression. We then computed the angle that explained tuning best by reducing the mean square error per hemisphere and bat (**Table S1**). The angle used for the rest of the analysis is the bat averaged angle 48º. The composite theoretical tonotopic gradient (**Figure 4D** and **Figure 4H**) was obtained using the averaged angle and averaged linear regression coefficients across bats.

Sound selectivity. An individual cell is referred to as selective if its response shows invariance to members of one category while showing different responses to a second category[30,41]. We computed the selectivity index for each cell as:

$$SI = \frac{1}{n}\sum_{i}^{n}\frac{(Xi - Yi)}{(Xi + Yi)}$$

with Xi as the median activity evoked within the response window (10 frames after sound onset, 312.5ms) by exemplar i of a stimulus type and Yi the corresponding activity evoked by an exemplar i of a second stimulus. Because multiple exemplars were presented, SI corresponds to the average of the selectivity index computed for pairs of exemplars. SI ranged from 1 to -1 with

15

361 positive SI indicating selectivity to the first category, while negative SI indicating selectivity to
362 the second category. Permutation tests were performed to assess cell selectivity significance (two-
363 sided, $\alpha<0.05$).
364

365 **Category decoding.** We evaluated the accuracy with which the population activity could be
366 classified as a natural stimulus exemplar by training and testing a linear discriminant classifier [42]
367 in a pairwise fashion, using a leave-one-out cross-validation. Two training population vectors (c_{s1},
368 c_{s2}) were obtained by averaging the activity for 320ms (10 frames) after stimulus onset for each
369 stimulus pair and each ROIs for a subset of randomly selected repetitions. The decoding vector w
370 was defined as:
371

372 $$w = c_{s1} - c_{s2}$$
373 and the bias as:

374 $$b = \frac{-(c_{s2} * w + c_{s1} * w)}{2}$$

375 The test population vector (remaining repetition, x) was classified as stimulus 1 or 2, following
376 rule:

377 $$y(x) = w * x + b$$
378 If $y(x)>0$, x is classified as stimulus 1 and as stimulus 2 if $y(x)<0$.
379 We repeated this procedure 100 times and averaged the accuracy computed for each stimulus pair
380 per bat. The accuracy presented in **Figure 3G** and **S5A** is averaged across bats. Similarly, for
381 **Figure 3F** we evaluated the accuracy with which the population could discriminate between the
382 two bat calls categories using all exemplars of each category per time point (corresponding to the
383 timing of a frame) and the same approach.
384 The weights presented in **Figure S5B-C** correspond to the average vector w over 100 cross-
385 validation.
386

387 **Basic auditory feature characterization.** To determine if basic functional auditory features could
388 underlie the population selectivity to natural bat calls, we compared a number of auditory
389 functional features for the selective and non-selective populations, including BF tuning, tuning
390 curve bandwidth, amplitude of evoke-response to various sounds, upsweed/downsweep tuning
391 (**Figure 3I-L** and **Figure S6**). Bandwidth is defined as the area under the curve for a normalized
392 tuning curve (small if the tuning is sharp and large if the tuning is broad). Population tuning to
393 downseep, upsweep rate and white noise duration was computed as the average of the tuning
394 curves normalized by the peak amplitude.
395

396 **Spatial clustering analysis** Local percentage of selective cells was computed in 30μm radial
397 increments for each selective cell for either social or navigation selectivity per recording sites.
398 Given the variations in site selectivity the local percentage was divided by the site overall
399 selectivity to compute a local selectivity ratio as a function of distance per bat. Spatial shuffle
400 distributions were computed per 30μm radial increments and imaging site using the x and y
401 coordinates (unpaired) to determine significance using a permutation test (two-sided, $\alpha<0.05$).
402 Selective clusters were identified on a composite map across sites and bats using Density-Based
403 Spatial Clustering of Applications with Noise (*DBSCAN*, based on the pairwise Euclidean distance
404 between ROIs). Clusters minimum number of cells was set to 3 and epsilon was set to 30μm.
405

References

1. May, B., Moody, D.B., and Stebbins, W.C. (1989). Categorical perception of conspecific communication sounds by Japanese macaques, Macaca fuscata. The Journal of the Acoustical Society of America *85*, 837–847. 10.1121/1.397555.

2. Barker, A.J., Veviurko, G., Bennett, N.C., Hart, D.W., Mograby, L., and Lewin, G.R. (2021). Cultural transmission of vocal dialect in the naked mole-rat. Science *371*, 503–507. 10.1126/science.abc6588.

3. Ehret, G. (2005). Infant Rodent Ultrasounds – A Gate to the Understanding of Sound Communication. Behav Genet *35*, 19–29. 10.1007/s10519-004-0853-8.

4. Wyttenbach, R.A., May, M.L., and Hoy, R.R. (1996). Categorical Perception of Sound Frequency by Crickets. Science *273*, 1542–1544. 10.1126/science.273.5281.1542.

5. Baugh, A.T., Akre, K.L., and Ryan, M.J. (2008). Categorical perception of a natural, multivariate signal: Mating call recognition in túngara frogs. Proceedings of the National Academy of Sciences *105*, 8985–8988. 10.1073/pnas.0802201105.

6. Reiss, D., McCowan, B., and Marino, L. (1997). Communicative and other cognitive characteristics of bottlenose dolphins. Trends in Cognitive Sciences *1*, 140–145. 10.1016/S1364-6613(97)01046-2.

7. Reichmuth, C., and Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. Current Opinion in Neurobiology *28*, 66–71. 10.1016/j.conb.2014.06.011.

8. Prather, J., Nowicki, S., Anderson, R., Peters, S., and Mooney, R. (2009). Neural Correlates of Categorical Perception in Learned Vocal Communication. Nat Neurosci *12*, 221–228. 10.1038/nn.2246.

9. Wright, G.S., Chiu, C., Xian, W., Wilkinson, G.S., and Moss, C.F. (2014). Social Calls Predict Foraging Success in Big Brown Bats. Current Biology *24*, 885–889. 10.1016/j.cub.2014.02.058.

10. Liu, S.T., Montes-Lourido, P., Wang, X., and Sadagopan, S. (2019). Optimal features for auditory categorization. Nat Commun *10*, 1302. 10.1038/s41467-019-09115-y.

11. Theunissen, F.E., and Elie, J.E. (2014). Neural processing of natural sounds. Nat Rev Neurosci *15*, 355–366. 10.1038/nrn3731.

12. Tsunada, J., and Cohen, Y.E. (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. Frontiers in Neuroscience *8*.

13. Yin, P., Strait, D.L., Radtke-Schuller, S., Fritz, J.B., and Shamma, S.A. (2020). Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex. Current Biology *30*, 1649-1663.e5. 10.1016/j.cub.2020.02.047.

17

14. Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci *15*, 536–548. 10.1038/nrn3747.

15. Surlykke, A., and Moss, C.F. (2000). Echolocation behavior of big brown bats, Eptesicus fuscus, in the field and the laboratory. The Journal of the Acoustical Society of America *108*, 2419–2429. 10.1121/1.1315295.

16. Gadziola, M.A., Grimsley, J.M.S., Faure, P.A., and Wenstrup, J.J. (2012). Social Vocalizations of Big Brown Bats Vary with Behavioral Context. PLoS One *7*, e44550. 10.1371/journal.pone.0044550.

17. Montoya, J., Lee, Y., and Salles, A. (2022). Social Communication in Big Brown Bats. Frontiers in Ecology and Evolution *10*.

18. Winer, J.A., and Schreiner, C.E. eds. (2005). The Inferior Colliculus (Springer-Verlag) 10.1007/b138578.

19. Barnstedt, O., Keating, P., Weissenberger, Y., King, A.J., and Dahmen, J.C. (2015). Functional Microarchitecture of the Mouse Dorsal Inferior Colliculus Revealed through In Vivo Two-Photon Calcium Imaging. J Neurosci *35*, 10927–10939. 10.1523/JNEUROSCI.0103-15.2015.

20. Winer, J.A., Larue, D.T., Diehl, J.J., and Hefti, B.J. (1998). Auditory cortical projections to the cat inferior colliculus. Journal of Comparative Neurology *400*, 147–174. 10.1002/(SICI)1096-9861(19981019)400:2<147::AID-CNE1>3.0.CO;2-9.

21. Guo, W., Chambers, A.R., Darrow, K.N., Hancock, K.E., Shinn-Cunningham, B.G., and Polley, D.B. (2012). Robustness of Cortical Topography across Fields, Laminae, Anesthetic States, and Neurophysiological Signal Types. J. Neurosci. *32*, 9159–9172. 10.1523/JNEUROSCI.0065-12.2012.

22. Schreiner, C.E., and Langner, G. (1997). Laminar fine structure of frequency organization in auditory midbrain. Nature *388*, 383–386. 10.1038/41106.

23. Klug, A., Bauer, E.E., Hanson, J.T., Hurley, L., Meitzen, J., and Pollak, G.D. (2002). Response Selectivity for Species-Specific Calls in the Inferior Colliculus of Mexican Free-Tailed Bats is Generated by Inhibition. Journal of Neurophysiology *88*, 1941–1954. 10.1152/jn.2002.88.4.1941.

24. Poon, P.W.F., Sun, X., Kamada, T., and Jen, P.H.-S. (1990). Frequency and space representation in the inferior colliculus of the FM bat, Eptesicus fuscus. Exp Brain Res *79*, 83–91. 10.1007/BF00228875.

25. Koay, G., Heffner, H.E., and Heffner, R.S. (1997). Audiogram of the big brown bat (Eptesicus fuscus). Hearing Research *105*, 202–210. 10.1016/S0378-5955(96)00208-0.

26. Wong, A.B., and Borst, J.G.G. (2019). Tonotopic and non-auditory organization of the mouse dorsal inferior colliculus revealed by two-photon imaging. eLife *8*, e49091. 10.7554/eLife.49091.

27. Oliver, D.L. (1987). Projections to the inferior colliculus from the anteroventral cochlear nucleus in the cat: Possible substrates for binaural interaction. Journal of Comparative Neurology *264*, 24–46. 10.1002/cne.902640104.

28. Wright, G., Chiu, C., Xian, W., Moss, C., and Wilkinson, G. (2013). Social calls of flying big brown bats (Eptesicus fuscus). Frontiers in Physiology *4*.

29. Chen, C., Xian, W., and Moss, C.F. (2008). Flying in silence: Echolocating bats cease vocalizing to avoid sonar jamming. Proceedings of the National Academy of Sciences of the United States of America *105*, 13116–13121. 10.1073/pnas.0804408105.

30. Salles, A., Park, S., Sundar, H., Macías, S., Elhilali, M., and Moss, C.F. (2020). Neural Response Selectivity to Natural Sounds in the Bat Midbrain. Neuroscience *434*, 200–211. 10.1016/j.neuroscience.2019.11.047.

31. Woolley, S.M.N., and Portfors, C.V. (2013). Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain. Hear Res *305*, 10.1016/j.heares.2013.05.005. 10.1016/j.heares.2013.05.005.

32. Kar, M., Pernia, M., Williams, K., Parida, S., Schneider, N.A., McAndrew, M., Kumbam, I., and Sadagopan, S. (2022). Vocalization categorization behavior explained by a feature-based auditory categorization model. eLife *11*, e78278. 10.7554/eLife.78278.

33. Achutha, A.C., Peremans, H., Firzlaff, U., and Vanderelst, D. (2021). Efficient encoding of spectrotemporal information for bat echolocation. PLOS Computational Biology *17*, e1009052. 10.1371/journal.pcbi.1009052.

34. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671–675. 10.1038/nmeth.2089.

35. Yang, G., Pan, F., Parkhurst, C.N., Grutzendler, J., and Gan, W.-B. (2010). Thinned-skull cranial window technique for long-term imaging of the cortex in live mice. Nat Protoc *5*, 201–208. 10.1038/nprot.2009.222.

36. Shih, A.Y., Mateo, C., Drew, P.J., Tsai, P.S., and Kleinfeld, D. (2012). A Polished and Reinforced Thinned-skull Window for Long-term Imaging of the Mouse Brain. J Vis Exp, 3742. 10.3791/3742.

37. Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L.F., Dalgleish, H., Carandini, M., and Harris, K.D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. 061507. 10.1101/061507.

510   38. Heckman, J.J., Proville, R., Heckman, G.J., Azarfar, A., Celikel, T., and Englitz, B. (2017).
511   High-precision spatial localization of mouse vocalizations during social interaction. Sci Rep
512   *7*, 3017. 10.1038/s41598-017-02954-z.

513   39. Wohlgemuth, M., and Moss, C.F. (2016). Midbrain auditory selectivity to natural sounds.
514   Proceedings of the National Academy of Sciences of the United States of America *113*,
515   2508–2513. 10.1073/pnas.1517451113.

516   40. Ito, T., and Oliver, D.L. (2012). The basic circuit of the IC: tectothalamic neurons with
517   different patterns of synaptic organization send different messages to the thalamus. Front
518   Neural Circuits *6*, 48. 10.3389/fncir.2012.00048.

519   41. Tsunada, J., Lee, J.H., and Cohen, Y.E. (2011). Representation of speech categories in the
520   primate auditory cortex. J Neurophysiol *105*, 2634–2646. 10.1152/jn.00037.2011.

521   42. Bagur, S., Averseng, M., Elgueda, D., David, S., Fritz, J., Yin, P., Shamma, S., Boubenec,
522   Y., and Ostojic, S. (2018). Go/No-Go task engagement enhances population representation
523   of target stimuli in primary auditory cortex. Nat Commun *9*, 2529. 10.1038/s41467-018-
524   04839-9.

525

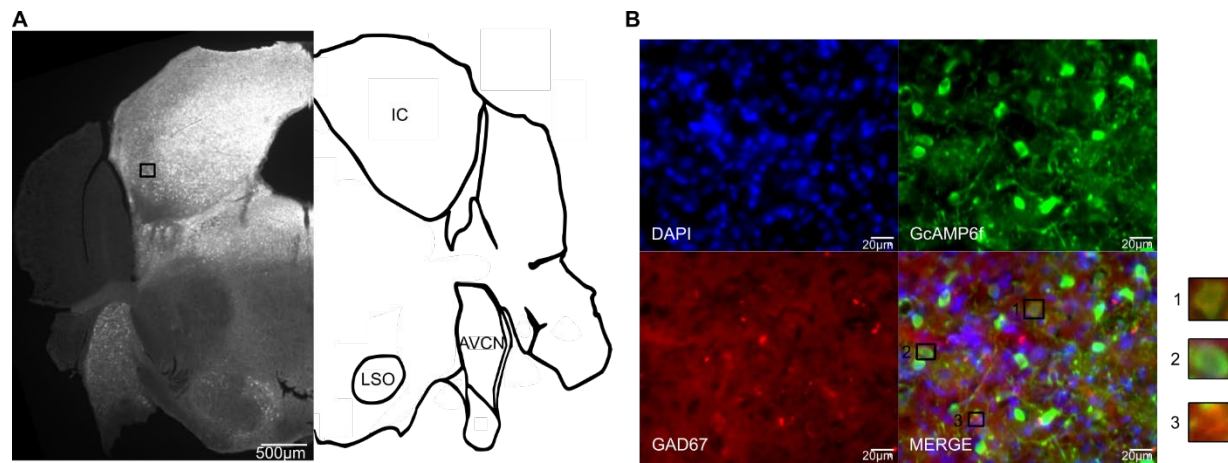526 **Supplementary Materials:**

527 Figures S1-7

528 Table S1

529

Figure S1. Histology and immunostaining for GCaMP6f and GAD67 A. GFP immunostaining shows expression of GCaMP6f throughout the IC. The black box indicates the location of the site highlighted in B. **B**. Immunostaining for IC locus highlighted in A for cell nucleus (DAPI, top left), GCaMP6f (GFP, top right), inhibitory neurons (GAD67, bottom left), and merged image (bottom right). Merged image suggests minimal overlap between GCcAMP6f and GAD67 expression. Example neurons numbered on the merged image are displayed on the right: 1) GAD67+ and GcAMP6f+ (yellow), 2) GcAMP6f+ and GAD67- (green), 3) GcAMP6f- and GAD67+ (red).

530
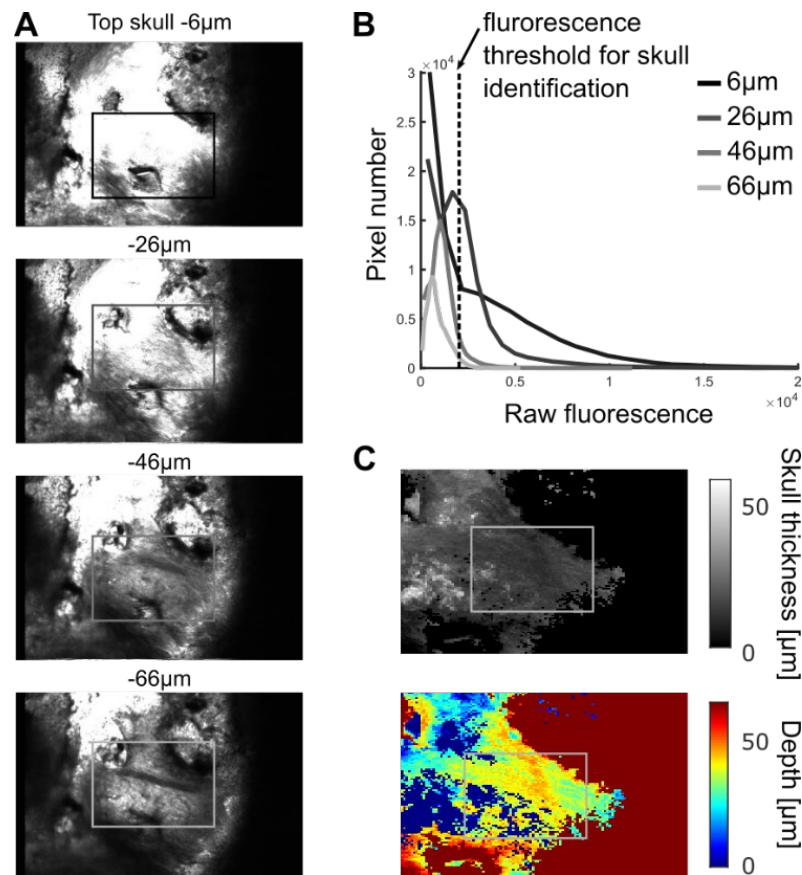
Figure S2. Site depth estimation corrected by skull thickness A. Mean images (150 frames) from 1X volumetric stack for example 2X imaging site (highlighted by the black to gray box, colors correspond to depth in B) every 10 μm until example site imaging depth (bottom image). **B**. Raw fluorescence distribution for images displayed in A. Note that shallower images possess a longer high fluorescence tail indicating a larger skull coverage. The criterion used for skull identification (95th percentile of the overall volumetric stack distribution) is displayed in the dashed line. **C.** Skull thickness and corresponding depth estimation for example volumetric stack in A and B.

531
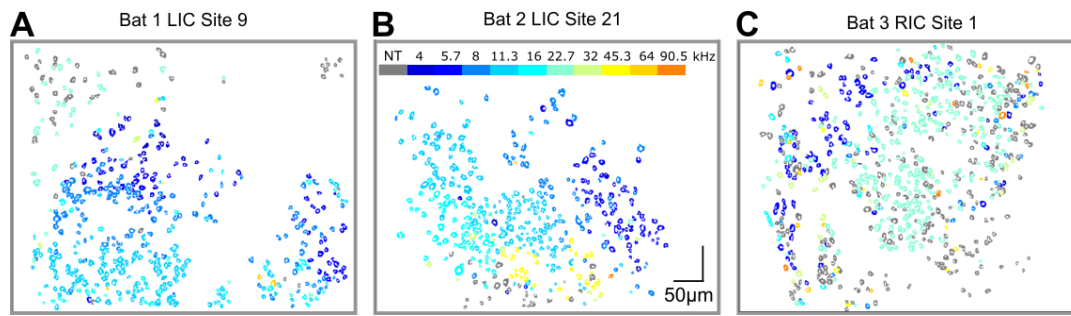
Figure S3. Tonotopic gradient in example sites for 3 separate bats. A. Example site from the left DCIC of bat 1. **B.** Example site from the left DCIC of bat 2. **C.** Example site from the right DCIC of bat 3.
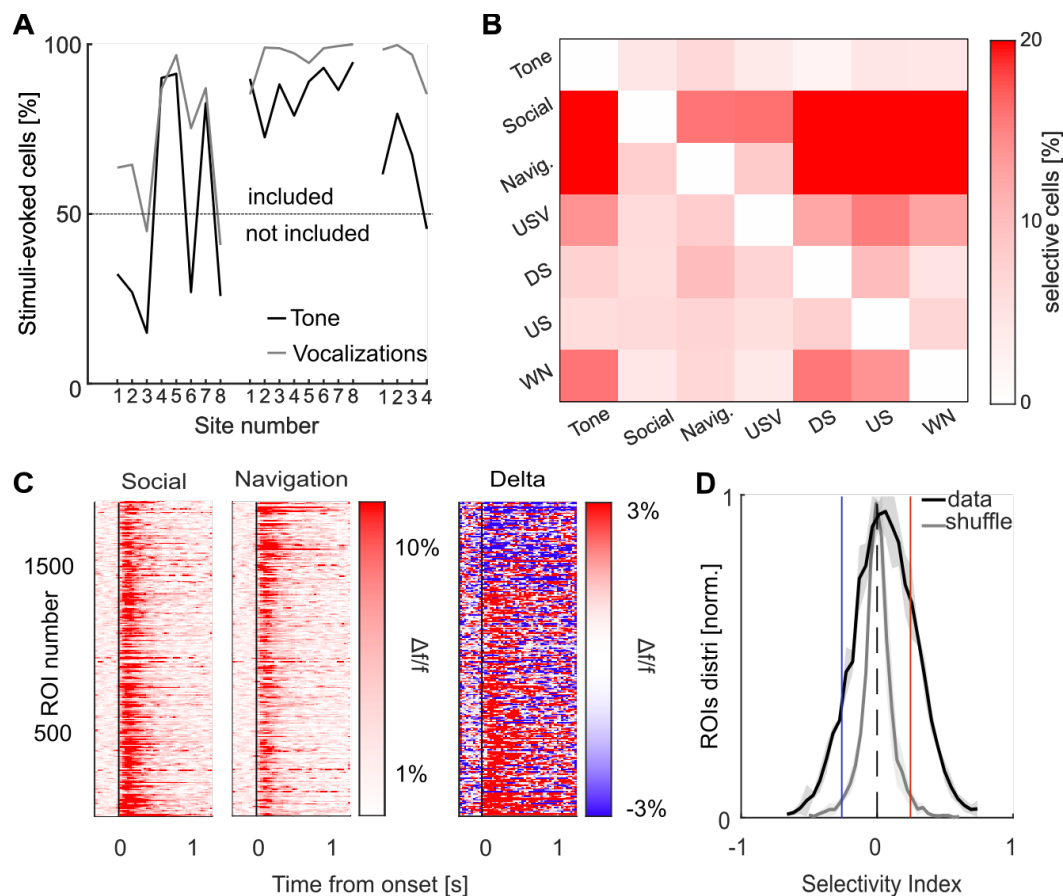
532

Figure S4. Selectivity index method. A. Percentage of cells showing significant stimulus-evoked responses (ANOVA, α<0.05) for tones (in black) and vocalizations (in gray) for each recorded site. Sites with fewer than 50% of cells with significant tone-evoked responses were excluded from the vocalization analysis. **B.** Mean percentage of significantly selective cells using SI for each sound category. Note that the percentage of selective cells to conspecific calls is consistently higher than other categories. **C.** Average sound-evoked Δf/f per category for significant selective cells sorted by their selectivity index (n_{cells} = 1885). Left: average sound-evoked activity for social calls. Middle: average sound-evoked activity for navigation calls. Right: difference between social and navigation activity. **D.** Average selectivity index distribution across bats (shading: SEM). Average significance boundaries are displayed in blue (2.5 percentile, SI_{bound} = -0.2959) and red (97.5 percentile, SI_{bound} = 0.3109).
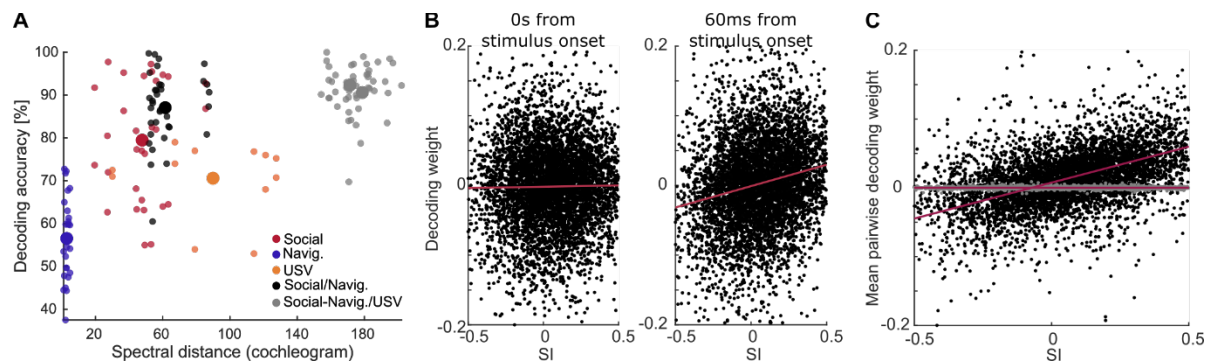
Figure S5. Linear decoding method A. Pairwise mean decoding accuracy as a function of spectral distance for social calls (in red), navigation calls (in blue), USVs (orange), conspecific calls (social vs navigation), or cross-species calls (social vs USVs, and navigation vs USVs). Smaller dots: individual exemplars pair values, larger dots: average for one comparison type. **B.** Decoding weights for timecourse decoder (Figure 3F) as a function of selectivity index per cell (black dot). *Left*: decoding weights does not increase (linear regression in red) with SI at 0s after stimulus onset. *Right*: decoding weights increases linearly with SI at 60ms after stimulus onset. **C.** Mean decoding weights from pairwise decoding for social vs navigation (Figure 3G-H) increase with SI (linear regression in red), but not for same stimulus decoding (social only, in grey)
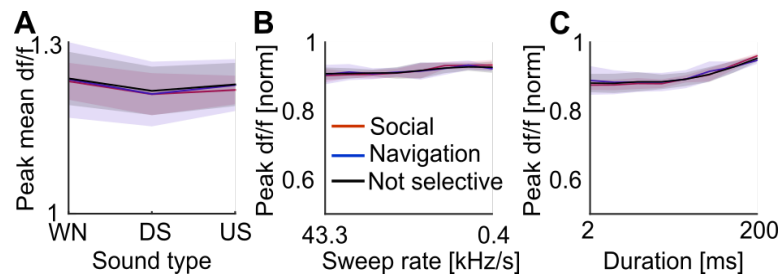
533

534

Figure S6. Additional auditory features for selective and non-selective populations A. Average peak evoked-responses as df/f to different type of complex sounds: white noise (WN), downsweep (DS) and upsweep (US), for social, navigation and non-selective cells (shading: SEM, ANOVA, p = 0.9721). **B.** Mean upsweep rate tuning as peak df/f normalized for social, navigation and non-selective cells (shading: SEM, ANOVA, p = 0.9907). **C.** Mean white noise duration tuning as peak df/f normalized for social, navigation and non-selective cells (shading: SEM, ANOVA, p =0.9648).
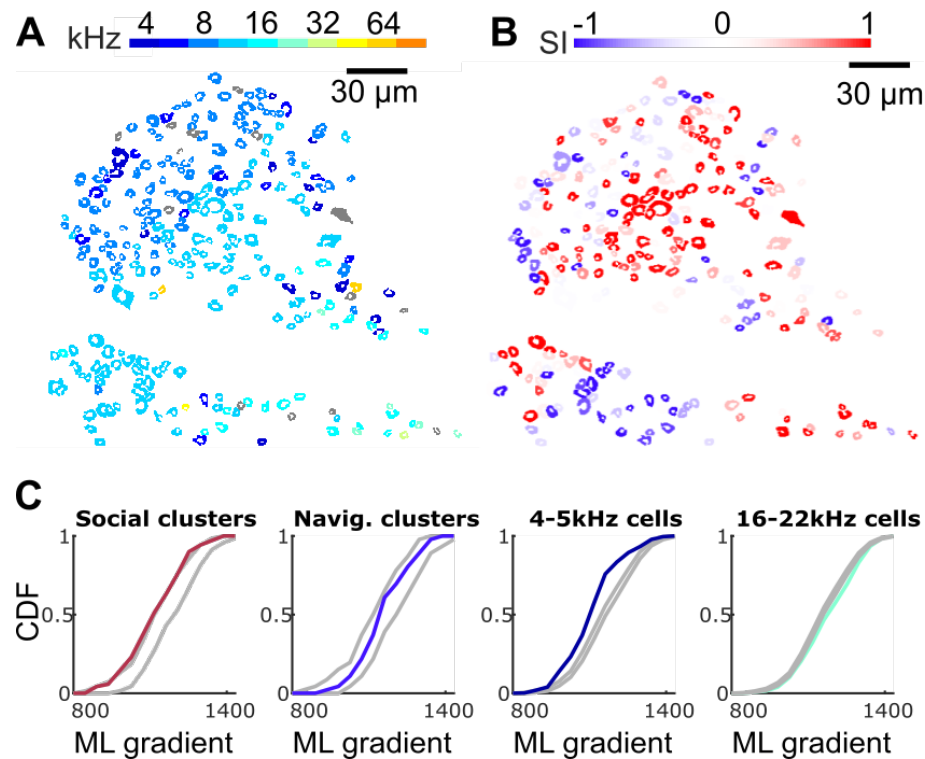
535

Figure S7. Relationship between tuning and selectivity index examples and composite view A. Tuning map for example site in Figure 4 color-coded according to the best frequency of the neuron. **B.** Corresponding social/navigation selectivity map for the same example site, color-coded by the neuron's selectivity index. Note that the selectivity 'hotspots' do not follow the tonotopic gradient. **C.** Cumulative distribution for cluster centers and corresponding shuffle distribution (randomly selected from all cells coordinates, in gray) along the composite RL gradient of all imaged sites. *Left*: social cluster falls just outside of the shuffle distribution (p<0.01), perhaps due to the increase in cluster size with cell density. *Center left*: navigation cluster centers fall within the 95% confidence interval therefore following cell density independently of the tonotopic gradient. *Center right*: cells tuned to low frequencies (4 and 5.7kHz, in dark blue) are located more rostromedially than the corresponding shuffled distribution (p<0.01). *Right*: cells tuned to high frequencies (16 and 22.7kHz) are located more caudolaterally than the corresponding shuffled distribution (p<0.01).

	Bat 1	**Bat 2**	**Bat 3**	**Sum**
Hemisphere	LIC	RIC, LIC	RIC	2 LIC, 2 RIC
Number of sites	8	5, 3	4	20
Tonotopy best fit angle	80	30,20	60	47.5
Tonotopy octave/mm	1.1	2.1, 3.5	1.5	2.1

Table S1. Tonotopic gradient angle estimation

536