

Noisy neuronal populations effectively encode sound localization in the dorsal inferior colliculus of awake mice

Juan C Boffi , Brice Bathellier, Hiroki Asari, Robert Prevedel 

Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany • Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory, Monterotondo, Italy • Institut Pasteur, Université Paris-Cité, INSERM, Institut de l'Audition, Paris, France • Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany • Molecular Medicine Partnership Unit, European Molecular Biology Laboratory, Heidelberg, Germany • Interdisciplinary Center for Neurosciences, Heidelberg University, Heidelberg, Germany

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

 A newer version is available. [Read the latest version.](#)

Reviewed Preprint

v3 • November 6, 2024

Reviewed Preprint

v2 • October 22, 2024

Revised by authors

Reviewed Preprint

v1 • June 17, 2024

eLife Assessment

The paper reports the **important** discovery that the mouse dorsal inferior colliculus, an auditory midbrain area, encodes sound location. The evidence supporting the claims is **solid**, being supported by both optical and electrophysiological recordings. The observations described should be of interest to auditory researchers studying the neural mechanisms of sound localization and the role of noise correlations in population coding.

<https://doi.org/10.7554/eLife.97598.2.sa4>

Abstract

Sound location coding has been extensively studied at the central nucleus of the mammalian inferior colliculus (CNIC), supporting a population code. However, this population code has not been extensively characterized on the single-trial level with simultaneous recordings or at other anatomical regions like the dorsal cortex of inferior colliculus (DCIC), which is relevant for learning-induced experience dependent plasticity. To address these knowledge gaps, here we made in two complementary ways large-scale recordings of DCIC populations from awake mice in response to sounds delivered from 13 different frontal horizontal locations (azimuths): volumetric two-photon calcium imaging with ~700 cells simultaneously recorded at a relatively low temporal resolution, and high-density single-unit extracellular recordings with ~20 cells simultaneously recorded at a high temporal resolution. Independent of the method, the recorded DCIC population responses revealed substantial trial-to-trial variation (neuronal noise) which was significantly correlated across pairs of neurons (noise correlations) in the passively listening condition. Nevertheless, decoding analysis supported that these noisy response patterns encode sound location on the single-trial basis, reaching errors that match the discrimination ability of mice. The detected noise correlations contributed to minimize the error of the DCIC population code of sound azimuth.

Altogether these findings point out that DCIC can encode sound location in a similar format to what has been proposed for CNIC, opening exciting questions about how noise correlations could shape this code in the context of cortico-collicular input and experience dependent plasticity.

Introduction

Locating a sound source facilitates essential behaviors such as foraging, mating and predator avoidance, thanks to the omni-directionality and long reach of the auditory system (King et al., 2001 [↗](#)). In vertebrates, sound localization relies on both binaural cues such as interaural level and time differences (ILD, ITD) (Knudsen and Konishi, 1979 [↗](#)) and monaural cues including spectral notches (SN) (Kulkarni and Colburn, 1998 [↗](#)). These sound localization cues are processed independently at brainstem nuclei in the ascending auditory pathway and integrated altogether for the first time at the inferior colliculus (IC) (Adams, 1979 [↗](#); Brunso-Bechtold et al., 1981 [↗](#); Felix et al., 2018 [↗](#); Grothe et al., 2010 [↗](#)). This makes the IC a crucial early relay of the ascending auditory pathway to study how a primary neural representation of auditory space is formed (Felix et al., 2018 [↗](#); Grothe et al., 2010 [↗](#)). Furthermore, the IC is also targeted by cortico-fugal interactions involved in higher order functions concerning sound localization information like experience dependent plasticity (Bajo et al., 2019 [↗](#), 2010 [↗](#); Bajo and King, 2012 [↗](#)), supporting that IC plays a complex part in the sound localization processing network, contributing from primary representation to shaping behavior.

Previous studies involving extracellular recordings from the mammalian IC revealed that average responses from IC neurons were tuned to multiple sound location cues, evidencing the integration of such cues at IC (Chase and Young, 2005 [↗](#)). However, the amount of sound location information carried by individual neurons throughout the auditory pathway is limited and quite variable, suggesting that the sound location information is distributed into a population code for auditory space (Clarey et al., 1995 [↗](#); Day and Delgutte, 2016 [↗](#); Groh et al., 2003 [↗](#); Panniello et al., 2018 [↗](#)). Fundamental work by Day and Delgutte (2013) [↗](#) supports that sound location in the horizontal plane (azimuth) is represented at the mammalian IC by the activity patterns of populations of neurons that display average response tuning to azimuth, while ruling out other possibilities such as a topological code, an opposing two channel model or a population vector code (Georgopoulos et al., 1986 [↗](#), 1982 [↗](#)). Nevertheless, this knowledge stems from prospective IC population activity patterns that were approximated by pooling non-simultaneous extracellular recordings and relied on extensive resampling of the pooled recordings to increase the number of “virtual” trials analyzed. Therefore, it remains an open question whether the actual activity patterns occurring in the IC of awake animals during single-trials can encode sound location given emerging properties such as correlated neuronal noise (Averbeck et al., 2006 [↗](#); Kohn et al., 2016 [↗](#); Sadeghi et al., 2019 [↗](#)). Trial-to-trial response variability (neuronal noise) and neuronal noise correlation can have a profound impact on the neural representations of sensory information, but this can only be determined through simultaneous large-scale recordings (Averbeck et al., 2006 [↗](#); Kohn et al., 2016 [↗](#)). Finally, traditional microelectrode recordings were mostly performed at a relatively deep anatomical subdivision of IC (central nucleus, CNIC), usually not targeting superficial regions (dorsal cortex, DCIC; external cortex, ECIC) (Chase and Young, 2008 [↗](#), 2005 [↗](#); Day and Delgutte, 2013 [↗](#); Guo et al., 2016 [↗](#); Lesica et al., 2010 [↗](#); Schnupp and King, 1997 [↗](#)). To overcome these limitations and explore the importance of trial-to-trial response variability, noise correlation and dorsal IC populations on sound location coding, we monitored DCIC population response patterns on the single-trial basis and interrogated the population code for sound location at the DCIC from passively listening awake mice.

Recent technical developments, including multichannel silicon probes (Jun et al., 2017 [↗](#)) and fast volumetric Ca^{2+} imaging (Prevedel et al., 2016 [↗](#); Stringer et al., 2019 [↗](#); Weisenburger et al., 2019 [↗](#)) enable routine simultaneous recordings from large numbers of neurons *in vivo*. Beyond

their high throughput, volumetric methods are also a promising approach to precisely interrogate the topological distribution of neurons sensitive to sound azimuth, which has been reported to be random at CNIC (Day and Delgutte, 2013), but is understudied at DCIC. In this work, we implemented scanned temporal focusing two-photon microscopy (sTeFo 2P) (Prevedel et al., 2016), an advanced volumetric Ca^{2+} imaging modality, to simultaneously record the activity of unprecedentedly large DCIC populations. Here we refer to DCIC as the dorsomedial IC region covering CNIC (Zhou and Shore, 2006). Our approach produced direct evidence supporting that the response patterns from mammalian DCIC populations effectively encode sound location on the single-trial basis in spite of their variability across trials. We also detected the occurrence of substantial noise correlations which can contribute to reducing the error of this population code. Furthermore, we complemented our imaging results with electrophysiological recordings with neuropixels probes (Jun et al., 2017), reaching conclusions that were generally aligned across imaging and electrophysiological experiments. Altogether, our findings point to a functional role of DCIC in sound location coding following a similar population coding mechanism to what has been proposed for CNIC (Day and Delgutte, 2013). While CNIC is the main ascending relay carrying sound location information from brainstem auditory nuclei to the cortex through the auditory thalamus (Adams, 1979; Brunso-Bechtold et al., 1981; Felix et al., 2018), DCIC is a higher order relay receiving profuse descending corticofugal inputs influencing auditory information processing including sound location (Bajo et al., 2019, 2010; Bajo and King, 2012; Lesicko et al., 2022; Winer et al., 2002). This knowledge sets forth exciting possibilities about the mechanisms by which cortico-collicular interactions involving DCIC and perhaps noise correlations affect relevant processes involving sound location information like experience dependent plasticity (Bajo et al., 2019, 2010; Bajo and King, 2012).

Results

Simultaneous recordings of DCIC population activity

Since the abundance and distribution of sound localization sensitive neurons at DCIC is not fully characterized, in a first instance we implemented sTeFo 2P for volumetric Ca^{2+} imaging to simultaneously record the activity from samples of DCIC neurons as large as technically possible. Moreover, TeFo based methods have been shown to be more resilient to highly scattering (optically opaque) tissues such as the IC (Dana and Shoham, 2012). Adeno-associated viral vector (AAV) transduced IC neurons expressing jRGECO1a (Fig. 1A) were imaged through a cranial window in awake head-fixed mice, passively listening to 500 ms-long broad-band noise stimuli (20–40 kHz band-passed white noise) delivered by a motorized speaker every 5 s located at one of 13 different frontal azimuth angles in a random order covering the frontal hemifield in 15° steps (Fig. 1B). We performed volumetric Ca^{2+} imaging (4 Hz volume rate) simultaneously sampling the activity of up to 2535 (643 ± 427 , median $\pm$ median absolute deviation, $n = 12$ mice) regions-of-interest (ROIs) that had a shape and size consistent with an IC neuronal soma (12–30 μm diameter, Suppl. Fig. 1A) (Schofield and Beebe, 2019), within a $470 \times 470 \times 500 \mu\text{m}$ volume from the right IC unilaterally, covering most of the dorso-ventral span of the anterior DCIC (Fig. 1A–C) and perhaps reaching the upper surface of CNIC below DCIC at more posterior locations (Paxinos and Franklin, 2001). Since sTeFo 2P trades off spatial resolution for acquisition speed to produce volume rates compatible with the kinetics of Ca^{2+} sensors (voxel size: $3.7 \times 3.7 \times 15 \mu\text{m}$) (Prevedel et al., 2016), we could not resolve fine spatial features of neurons such as dendrites or synaptic structures like spines in our images to unequivocally identify the segmented ROIs as individual neurons (Suppl. Fig. 1A). For this reason, we refer to these ROIs as “Units”, in an analogous fashion to extracellular electrode recordings studies. To compensate for the relatively slow decay time of the Ca^{2+} indicator signal and non-linearities in the summation of Ca^{2+} events, we evaluated the spike probabilities associated to the neuronal Ca^{2+} signals (Deneux et al., 2016) (Fig. 1D, Suppl. Fig. 1B). The spike probability estimation does not involve predicting precise spike timings (Deneux et al., 2016; Huang et al., 2021; Pachitariu et al., 2018; Rupprecht et al.,

2021 [Chase and Young, 2008](#)). Hence, we do not make any observation or interpretation of spike probabilities in relation to the fine temporal structure of neuronal spiking. Instead, we treated the spike probabilities as a measure proportional to the spike count evaluated within the time interval set by our volumetric imaging rate (4 Hz). Taking into account that the main carrier of sound location information is IC neuron firing rate ([Chase and Young, 2008](#)), we defined the evoked DCIC population responses as the total (summed) spike probability of each imaged unit during the sound stimulation epoch (500 ms). In this way, we systematically collected the simultaneous response patterns of DCIC populations to every azimuth trial.

The DCIC populations simultaneously imaged from passively listening awake mice displayed spontaneous activity that was not time-locked to the stimulation (on-going activity) and sound-evoked response patterns that varied markedly across trials, as did the overall, average response of the complete imaged population (**Fig. 1E**). By monitoring face movements videographically, we observed that the overall average population activity shows significant correlation to facial movements, specifically of the snout region (**Suppl. Fig 2A-C**). On the other hand, the correlation between the recorded activity from the individual units imaged and face movements was generally low with narrow distributions centered at 0 (**Suppl. Fig 2D**), pointing out that correlation to face movement is more evident at the population level.

To further corroborate our volumetric imaging data, we recorded DCIC population activity electrophysiologically through single-unit recordings with neuropixels probes in passively listening mice stimulated with 200 ms-long broad-band noise stimuli (15-50 kHz band-passed white noise) delivered by a motorized speaker every 3 s from 13 different frontal azimuth angles in a random order covering the frontal hemifield in 15° steps (**Fig. 1F-H**). The higher time resolution from the electrophysiological approach (30kHz acquisition rate) allowed us to deliver more frequent trials with shorter stimuli to effectively collect more trials following the same head fixation paradigm as the one used for imaging (see methods). The electrophysiology setup also produced lower background noise sound pressure level (SPL, 35.96 dB R.M.S.) in comparison to the sTeFo scope (44.83 dB R.M.S.), which enabled the use of a broader band stimulus favoring sound location cues at appropriate levels above background noise (at least 10 dB R.M.S., see methods). We simultaneously recorded up to 43 (21 ± 9 , median $\pm$ median absolute deviation, $n = 4$ mice) spike sorted and manually curated DCIC single-units (**Suppl. Fig. 1C-E**). We defined sound evoked responses from electrophysiologically recorded single-units as the spike count observed during sound stimulus presentation ([Chase and Young, 2008](#); [Day and Delgutte, 2016](#), [2013](#)). We determined which recording channels from the probe were located within DCIC both histologically with respect to the reference atlas ([Paxinos and Franklin, 2001](#)) (**Fig. 1F**) and functionally by analyzing the arrangement of sound frequency sensitive units along the shank of the probe. We detected sound frequency sensitive single-units as units that displayed significant response dependency to sound frequency (pure tone stimuli) through χ^2 tests. The detected frequency dependent single-units displayed Gaussian tuning with clear peak best frequencies (**Suppl. Fig. 1F, G**). The observed relationship between frequency dependent DCIC unit best frequency and recording depth was not tonotopically arranged like in CNIC, which is consistent with DCIC location ([Barnstedt et al., 2015](#); [Wong and Borst, 2019](#)) (**Suppl. Fig. 1H**). The simultaneous, electrophysiologically recorded DCIC population activity from awake passively listening mice also displayed spontaneous on-going activity, which reflected in the overall average response of the complete recorded population, and sound-evoked response patterns that varied markedly across trials (**Fig. 1I**). In this section we used the word “noise” to refer to the sound stimuli used, the recording setup background sound levels or recording noise in the acquired signals. To avoid confusion, from now on the word “noise” will be used in the context of neuronal noise, which is the trial-to-trial variation in neuronal responses unrelated to stimuli, unless otherwise noted.

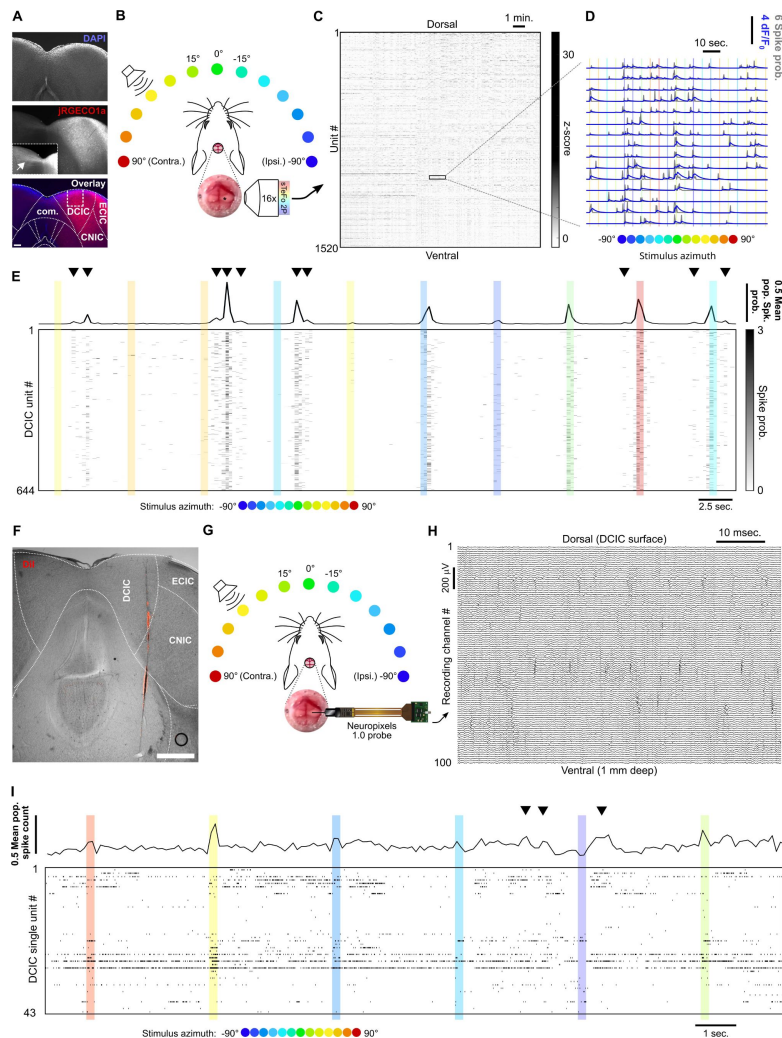


Figure 1

Simultaneous recording of DCIC population responses to sound azimuth through sTeFo-2P Ca^{2+} imaging and neuropixels probes.

A) Representative histological section showing AAV transduced jRGECO1a expression across IC. Middle panel inset: Contrast enhanced commissure region of the slice to visualize commissural projections from jRGECO1a expressing IC neurons. Bottom panel: Dotted lines delimit anatomical IC regions according to Paxinos et al. (2001); dashed lines delimit approximate area targeted for imaging. Scale bar: 200 μm . DCIC: dorsal cortex from inferior colliculus. CNIC: Central nucleus from inferior colliculus. ECIC: External cortex from inferior colliculus. Com.: Commissure from inferior colliculus. B) Schematic representation of the experimental design, incorporating sTeFo 2P for Ca^{2+} imaging. C) Neuropil-corrected and denoised jRGECO1a signals extracted from a representative full dataset. Extracted signals are arranged from dorsal (top) to ventral (bottom) ROI position across the DCIC volume imaged. D) Representative neuropil corrected and denoised jRGECO1a traces (blue) with their corresponding spike probability traces (gray) and stimulation epochs (color-coded based on stimulus azimuth angle according to (B)) super-imposed. E) Representative simultaneous recording of a DCIC population from an awake, passively listening mouse, displaying spontaneous, on-going activity (not synchronized to stimulation, arrowheads) and variable sound-evoked response patterns (during sound stimuli). Top trace is the population average response. Sound stimulation epochs are color-coded based on azimuth. F) Representative histological section showing DiI labeled neuropixels electrode tract across IC. Dotted lines delimit anatomical IC regions according to Paxinos et al. (2001). Scale bar: 500 μm . G) Same as (B) but representing integration with electrophysiological recording of DCIC population activity with a neuropixels probe. H) Representative high pass filtered (> 300 Hz) voltage traces simultaneously recorded from 100 channels spanning across DCIC in a neuropixels probe shank during an experiment displaying clear unit waveforms captured across neighbouring channels. I) Same as (E) but showing a representative raster plot of the spike sorted DCIC single-unit activity simultaneously recorded during an experiment.

Decoding sound azimuth from single-trial DCIC population responses

The observed variability in our imaging and electrophysiological data raised the following question: To what extent do the single-trial responses of the DCIC units carry information about stimulus azimuth? Since the number of DCIC neurons needed to effectively encode sound azimuth is not known, we first evaluated how accurately stimulus azimuth can be predicted from the high throughput single-trial DCIC population response patterns volumetrically imaged, without taking resampling strategies or pooling recorded responses across animals or recording sessions as in previous electrophysiological and imaging studies (Day and Delgutte, 2013 [↗](#); Jazayeri and Movshon, 2006 [↗](#); Panniello et al., 2018 [↗](#)). We performed cross-validated, multi-class classification of the single-trial population responses (decoding, **Fig. 2A** [↗](#)) using a naive Bayes classifier to evaluate the prediction errors as the absolute difference between the stimulus azimuth and the predicted azimuth (**Fig. 2A** [↗](#)). We chose this classification algorithm over others due to its generally good performance with limited available data. We visualized the cross-validated prediction error distribution in cumulative plots where the observed prediction errors were compared to the distribution of errors for random azimuth sampling (**Fig. 2B** [↗](#)). When decoding all simultaneously recorded units, the observed classifier output was not significantly better (shifted towards smaller prediction errors) than the chance level distribution (**Fig. 2B** [↗](#)). The classifier also failed to decode complete DCIC population responses recorded with neuropixels probes (**Fig. 3A** [↗](#)). Given the high dimensionality of our datasets (number of simultaneously recorded units) and the relatively low number of azimuth presentations collected (10-20 repetitions of each), we reasoned that the classification failure could be due to overfitting rather than to a lack of azimuth information in the dataset. Thus, we tested if reducing the dimension of the dataset helps avoid overfitting. Specifically, we used the same decoder after dimensionality reduction through principal component analysis (PCA) with both PCA and classifier fit being cross-validated (**Fig. 2C** [↗](#)). For both our imaging and electrophysiological datasets, we observed that decoding based on subsets of first principal components (PCs) produced absolute cross-validated single-trial prediction error distributions significantly better than the chance level distribution (**Fig. 2C** D, E, **Fig. 3** [↗](#) B-D); and that using larger numbers of PCs for model fitting and decoding decreased the performance (**Fig. 2D** [↗](#), E, **Fig. 3C** [↗](#), D). Altogether, these results support that DCIC population responses indeed carry stimulus azimuth information which can be effectively decoded on a single-trial basis; and that the failure of decoding from the full population was merely due to overfitting.

Sound azimuth information is carried redundantly by specific DCIC units

The decoded PCs consist of linear combinations of the responses from the units in the imaged or recorded DCIC population. Then the following questions arise: To what extent is the sound azimuth information distributed across the DCIC populations? Is it fully distributed across most units or specific to some of them, such as azimuth tuned neurons?

Since the abundance or distribution of azimuth sensitive units across DCIC is not fully known, we firstly asked if any fraction of the volumetrically imaged DCIC units carried more azimuth information in their responses than other simultaneously imaged units in the volume. We calculated the signal-to-noise ratio for the responses of each imaged unit (“neuronal” S/N, nS/N), defined as the ratio between the mean (signal) and the standard deviation (noise) of the responses to the azimuth trials evoking maximal responses (best azimuth). Imaged DCIC units generally showed a relatively low nS/N during sound stimulations (< 1 , **Fig. 4A** [↗](#)). Nevertheless, this nS/N was significantly larger than the one registered in the absence of sound stimulation (on-going activity from the periods between sound stimuli, **Fig. 1D** [↗](#), 4A), indicating that sound responses

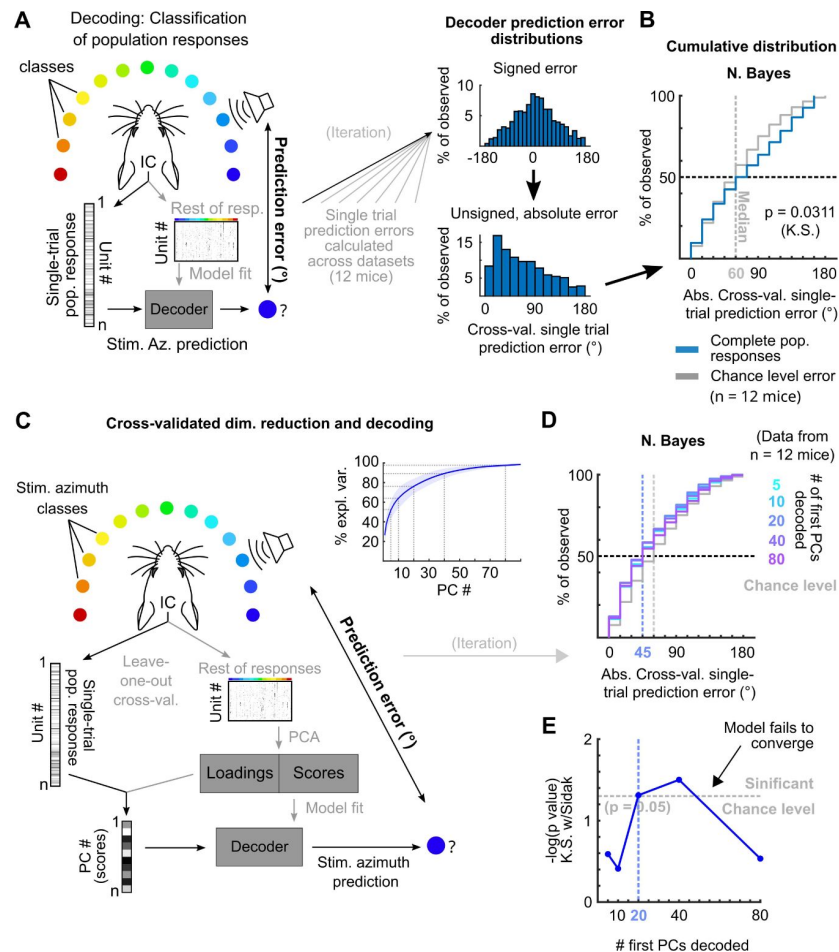


Figure 2

Decoding single-trial DCIC population responses to sound azimuth.

A) Schematic representation of the decoding strategy using multi-class classification on the recorded simultaneous population responses. B) Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained using naive Bayes classification (blue) and chance level distribution associated with our stimulation paradigm obtained by considering all possible prediction errors for the 13 azimuths tested (gray). K.S.: Kolmogorov-Smirnov test, n.s.: ($p > 0.05$). C) Schematic representation of the decoding strategy using the first PCs of the recorded population responses. Inset: % of explained variance obtained using PCA for dimensionality reduction on the complete population responses. Median (blue line) and median absolute deviation (shaded blue area) are plotted for ($n = 12$ mice/imaging sessions). D) Same as (B) but for decoding different numbers of first PCs from the recorded complete population responses. E) Significance of classification performance with respect to chance level for different numbers of first PCs, determined via Kolmogorov-Smirnov tests with Sidak correction for multiple comparisons. Arrowhead indicates model loss of performance associated with fitting more parameters for a larger feature space (# PCs) with the same dataset size (# trials collected).

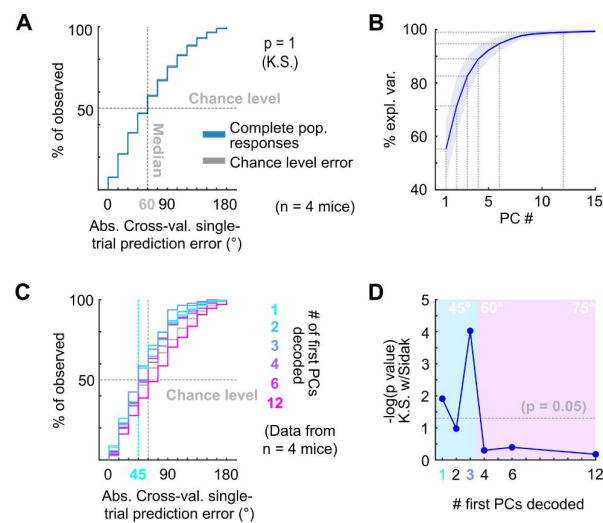


Figure 3

A) Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained decoding the complete simultaneously recorded population responses with neuropixels probes across mice (blue) and chance level distribution associated with our stimulation paradigm (gray). K.S.: Kolmogorov-Smirnov test, n.s.: ($p > 0.05$). B) % of observed variance explained across PC number for the complete population responses recorded with neuropixels. Median (blue line) and median absolute deviation (shaded blue area) are plotted for $n = 4$ mice. C) Same as (A) but for decoding different numbers of first PCs. D) Significance of decoding performance shown in (C) with respect to chance level for different numbers of first PCs, determined via Kolmogorov-Smirnov tests with Sidak correction for multiple comparisons. Shaded areas show the corresponding median decoding errors to the points within the area.

across the sampled units were often noisy. However, the broad distribution of nS/N also suggested that a sparse population of neurons responded more robustly. Next, we evaluated azimuth tuning to delimit subpopulations with higher azimuth selectivity. Using non-parametric one-way ANOVA (Kruskal-Wallis test), we identified the imaged units whose median responses changed significantly based on stimulus azimuth, which we defined as azimuth tuned units (**Fig. 4B**). The units with significant azimuth tuning represented only $8 \pm 3\%$ (median $\pm$ median absolute deviation, $n = 12$ mice) of the DCIC units simultaneously-imaged in one session (mouse; **Fig. 4C**). This is not significantly different from the false positive detection rate of our ANOVA tests ($\alpha = 0.05$, **Fig. 4C**). Given that a large number of trials showed no response at all in many units, it is possible that the large trial-to-trial variability in the imaged DCIC responses made it difficult to detect response specificity based on differences in median activity. We thus employed more sensitive χ^2 tests to determine the statistical dependence between the recorded DCIC responses and presented stimulus azimuth. We then ranked the units in our simultaneously imaged samples based on the p-value of the χ^2 tests (**Fig. 4D**) to delimit subsamples of DCIC units showing the strongest azimuth-dependent responses. The units with significant response dependency to stimulus azimuth ($p < 0.05$) represented $32 \pm 6\%$ (median $\pm$ median absolute deviation, $n = 12$ mice) of the units simultaneously-imaged in one session (mouse, **Fig. 4E**). Even the top ranked units typically did not respond in many trials; nevertheless, their evoked responses showed a tendency to be more selective towards contralateral or central stimulus azimuths (**Fig. 4D** left inset). The nS/N of top ranked units was also significantly larger than the one registered in the absence of sound stimulation, but these units did not display a major improvement in nS/N with respect to the complete population of imaged units, indicating that despite their response dependency to stimulus azimuth their responses are also noisy (**Fig. 4A, D** right inset).

To test if the single-trial response patterns from simultaneously imaged units with top response dependency to stimulus azimuth carry enough information to generate better azimuth predictions than the ones observed with the complete samples (**Fig. 2**), we evaluated the prediction errors obtained by decoding them. In doing so, we cross-validated both unit ranking based on response dependency (feature selection) and decoder fit (**Fig. 4F**). From this point on, we will refer to groups of units showing maximum response dependency to stimulus azimuth as “top ranked units”. To evaluate the minimum number of top ranked units necessary to generate stimulus azimuth predictions significantly better than the chance level, we decoded different sized subsamples of these units. We found that the single-trial response patterns of at least the 7 top ranked units produced stimulus azimuth prediction errors that were significantly smaller than chance level ($p = 6 \times 10^{-4}$, Kolmogorov-Smirnov with Sidak) using Bayes classification (naive approximation, for computation efficiency) (**Fig 4G**). Increasing the number of top ranked units decoded did not bring major improvements in decoder performance (plateau), up to a point where performance dropped, due to classifier overfitting likely caused by the relatively small number of collected trials per class (**Fig 4G**, bottom). If we consider the median of the prediction error distribution as an overall measure of decoding performance, the single-trial response patterns from subsamples of at least the 7 top ranked units produced median decoding errors that coincidentally matched the reported azimuth discrimination ability of mice (**Fig 4G**, minimum audible angle = 31°) (Lauer et al., 2011).

To further characterize the identified top ranked units, we studied how they are distributed across the imaged DCIC volume, which was positioned roughly at the center from the surface of the mouse dorsal IC (**Fig. 1A, B**). By comparing the positions of top ranked units to those of the complete samples of simultaneously imaged units across mice, we observed that the top ranked units ($\sim 32\%$ from the complete DCIC populations imaged) are scattered across the imaged volumes following the same distributions across the anatomical axes of the complete imaged populations (**Suppl. Fig. 3**). This observation suggests that the subpopulations of top ranked DCIC units associated with the population code of sound azimuth scatter across the DCIC without following a specific spatial pattern. We observed a tendency for a small shift towards positive

correlations in the correlation coefficient distribution between snout movement and the imaged activity of individual top ranked units, however this distribution was narrow and centered close to 0 (Suppl. Fig. 2E [↗](#)).

The observed broad distribution across DCIC and the relatively small number of top ranked units necessary to decode stimulus azimuth supports the characterization of this subpopulation with neuropixels data. Top ranked units detected in our neuropixels datasets also displayed response selectivity towards contralateral and central azimuths and did not respond in many trials (Fig. 5 [↗](#) A, B). The nS/N of the DCIC single-units recorded with neuropixels was again significantly larger than the one registered without sound stimulation, with a distribution tailing above 1. Top ranked units displayed a further shift towards higher nS/N with respect to the complete population of neuropixels recorded DCIC single-units. Still the medians of these nS/N distributions was below 1, supporting that DCIC single-unit responses are noisy (Fig. 5C [↗](#)). Top ranked units detected with neuropixels recordings represented $40 \pm 2\%$ (median $\pm$ median absolute deviation, $n = 4$ mice) of the units simultaneously-recorded in an experiment (mouse, Fig. 5D [↗](#)).

Interestingly, the nS/N distributions of neuropixels recorded DCIC single-units in response to pure tone stimuli were similar to those observed for broadband stimulus azimuth (Suppl. Fig. 11 [↗](#), Fig. 5C [↗](#)), suggesting that the observed nS/N levels in our imaging and electrophysiological recordings might be a property of the DCIC network (and not due to the recording method's sensitivity) which is noisy independently of the stimulus presented, at least in the passively listening condition. The percentage of sound frequency dependent units was $42 \pm 15\%$ (median $\pm$ median absolute deviation, $n = 4$ mice, Fig. 5D [↗](#)). Out of the DCIC single-units from our neuropixels recordings, $19 \pm 6\%$ (median $\pm$ median absolute deviation, $n = 4$ mice, Fig. 5D [↗](#)) displayed significant response dependency to both broadband stimulus azimuth and pure tone stimulus frequency, pointing out that not all DCIC azimuth dependent units are sensitive to sound frequency and vice-versa. To explore a possible relationship between the best frequency and azimuth sensitivity of these units in the context of the duplex theory (Rayleigh, 1907 [↗](#)), we evaluated the correlation between their best frequency and the significance level of their response dependency to azimuth ($-\log(p \text{ value})$ of the χ^2 test). This correlation was low ($R = 0.03$, 17 single-units from 4 mice), but we could observe that the clear majority of these units had best frequencies above $\sim 10\text{kHz}$, where the main sound location cues used by mice are carried (Lauer et al., 2011 [↗](#)) (Fig. 5E).

Decoding analysis (Fig. 4F [↗](#)) of the population response patterns from azimuth dependent top ranked units simultaneously recorded with neuropixels probes showed that the 4 top ranked units are the smallest subsample necessary to produce a significant decoding performance that coincidentally matches the discrimination ability of mice (31° (Lauer et al., 2011 [↗](#))) (Fig. 5F, G [↗](#)). Altogether, the close resemblance of the results obtained through volumetric imaging and electrophysiologically with neuropixels support that, even though noisy DCIC single-units can encode sound location individually with low performance (Fig. 5F, G [↗](#), first top ranked unit), a population code consisting of the simultaneous response patterns from small subsets of top ranked units occurs at DCIC, achieving more effective encoding errors.

Noise correlations and their contribution to the population code of sound azimuth at DCIC

We next investigated the occurrence of correlated variation in the trial-to-trial responses to each stimulus azimuth from pairs of simultaneously imaged or electrophysiologically recorded top ranked units as the occurrence of such neuronal noise correlations, which have been previously reported to occur at mammalian IC populations (Sadeghi et al., 2019 [↗](#)), can influence the accuracy of the DCIC population code for sound azimuth (Averbeck et al., 2006 [↗](#); Kohn et al., 2016 [↗](#)) (Fig. 6A [↗](#)). We evaluated the Kendall Tau as an unbiased, non-parametric pairwise correlation coefficient for the neuronal noise from all possible pairs of simultaneously recorded DCIC units across trial repetitions for each stimulus azimuth. Using cross-validated hierarchical

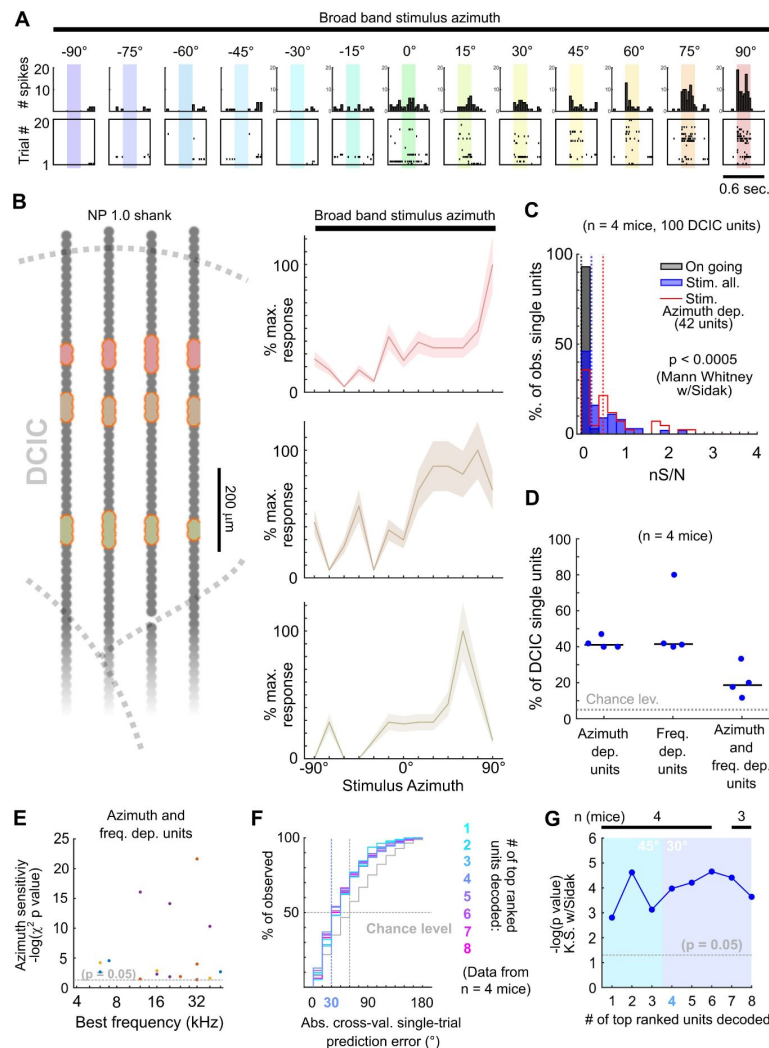


Figure 5

Neuropixels recordings support observations drawn from sTeFo 2P Ca^{2+} imaging experiments.

A) Representative responses to stimulus azimuth of a top ranked unit recorded with neuropixels. Top panels show the peri-stimulus time histograms and the bottom panels show the corresponding spike raster plots across trials. B) Left: Schematic representation of a neuropixels probe shank highlighting in different colors the position from channels where representative top ranked single-units were detected across DCIC (approximated histologically, same units as displayed in **Suppl. Fig 1C-E**). Right: Representative azimuth tuning curves from three DCIC top ranked single-units recorded with neuropixels, plot colors correspond to position in the shank schematic. Mean and standard deviation are plotted. C) Neuronal signal to noise level (nS/N) histograms from neuropixels recorded DCIC single-units in the absence of sound stimuli (on going activity, gray) and in response to the best azimuth trials, for all collected single-units (blue) and the top ranked single-unit subset (red). D) Percentages of sound azimuth dependent units (top ranked units), sound frequency dependent units and both azimuth and frequency dependent units across mice. Median value across mice is represented by a horizontal line. E) Relationship between azimuth sensitivity and best frequency of DCIC sound frequency and azimuth dependent single-units across mice. Data from n = 4 mice, point color corresponds to the same mouse. X axis scale is logarithmic. F) Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained by decoding the responses from different numbers of top ranked units simultaneously recorded with neuropixels probes across mice and chance level distribution associated with our stimulation paradigm (gray). G) Significance of decoding performance shown in (F) with respect to chance level for different numbers of top ranked units decoded, determined via Kolmogorov-Smirnov tests with Sidak correction for multiple comparisons. Shaded areas show the corresponding median decoding errors to the points within the area. Sample sizes (number of mice) is informed at the top of the graph for each point.

clustering of the simultaneously imaged or neuropixels recorded top ranked units based on noise correlation, we observed no clear groups (clusters, subpopulations) of highly noise correlated units across the trial repetitions from each stimulus azimuth (**Fig. 6B, C**). However, the pairwise noise correlation coefficients observed across datasets (mice) showed a distribution that was significantly shifted towards positive values with respect to chance level, calculated from the same datasets subjected to randomization of each unit's responses across trial repetitions for each stimulus azimuth (decorrelated, **Fig. 6A-C**). Nevertheless, this shift was much smaller for ipsilateral and central azimuths in our neuropixels data (**Fig. 6B, C**). Altogether these observations suggest that pairs of the DCIC top ranked units associated with the population code for sound azimuth display positive noise correlations with a likelihood that is higher than chance.

To characterize how the observed positive noise correlations could affect the representation of stimulus azimuth by DCIC top ranked unit population responses, we compared the decoding performance obtained by classifying the single-trial response patterns from top ranked units in the modeled decorrelated datasets versus the acquired data (with noise correlations). With the intention to characterize this with a conservative approach that would be less likely to find a contribution of noise correlations as it assumes response independence, we relied on the naive Bayes classifier for decoding throughout the study. Using this classifier, we observed that the modeled decorrelated datasets produced stimulus azimuth prediction error distributions that were significantly shifted towards higher decoding errors (**Fig. 6B, C**) and, in our imaging datasets, were not significantly different from chance level (**Fig. 6B**). Altogether, these results suggest that the detected noise correlations in our simultaneously acquired datasets can help reduce the error of the IC population code for sound azimuth.

Discussion

By performing fast volumetric Ca^{2+} imaging through sTeFo 2P microscopy (Prevedel et al., 2016) and single-unit recordings with neuropixel probes (Jun et al., 2017), here we tackled the technical challenge to simultaneously record the activity of a large number of DCIC units in response to sound azimuth. We show that sTeFo 2P effectively achieved high throughput sampling across large imaging depths in a highly light-scattering tissue such as the mouse IC (**Fig. 1**), showcasing the capability of sTeFo 2P for interrogating neuronal population activity over large volumes. Despite the advantages of volumetric sTeFo 2P, it also has limitations. In particular, large-scale volumetric imaging has a low temporal resolution, reaching sampling rates of 4 volumes per second, and requires the use of an indirect method of monitoring neuronal activity via Ca^{2+} sensors. The low temporal resolution effectively produces a low-pass filtering of neuronal activity, misrepresenting peak responses. Furthermore, indirectly inferring neuronal spiking responses from Ca^{2+} sensor signals can cause further information loss. Nevertheless, the main observations drawn from sTeFo 2P imaged population response patterns could be validated via single-unit recordings with new generation multichannel silicon probes (neuropixels), which have excellent temporal resolution but markedly lower throughput. This supports that the sTeFo imaging datasets still carried sufficient information about stimulus azimuth (**Fig. 2**, 4), showcasing how sTeFo 2P opens new possibilities for future studies requiring high-throughput recordings of simultaneous population activity, especially considering the advantages brought by further technical improvements such as multiplexing or larger fields of view (Clough et al., 2021; Demas et al., 2021; Ota et al., 2021; Weisenburger et al., 2019) and 3P excitation (Weisenburger et al., 2017).

The imaged or electrophysiologically recorded population activity datasets here described support with simultaneous recordings that the single-trial response patterns from subsets of neurons with response dependency to stimulus azimuth (top ranked units) constitute a population code for sound azimuth at the DCIC of awake, passively listening mice (**Fig. 4-5**). This finding complements previous studies analyzing pooled, non-simultaneous extracellular recordings from

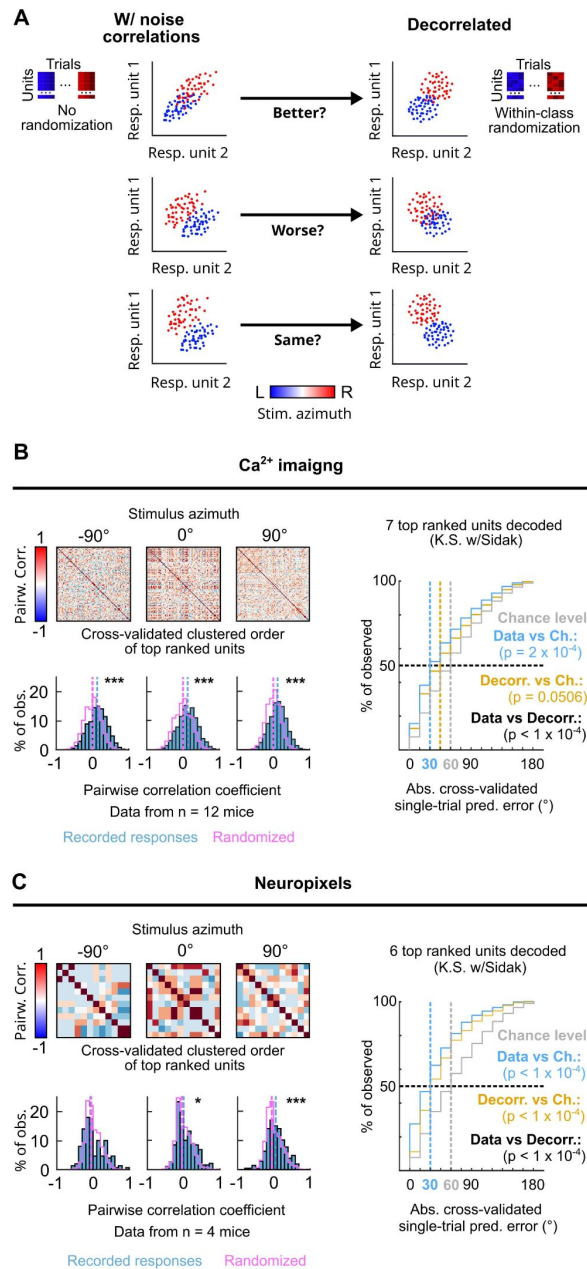


Figure 6

Noise correlations in DCIC population activity contribute to encode sound azimuth.

A) Simplified schematic representation of the possible effects from (positive) noise correlations on the response separability of a theoretical population consisting of 2 units, and within class randomization strategy to model decorrelated datasets lacking noise correlations. B, C) Left top: Representative correlation matrices of pairwise correlations between the responses of top ranked units detected in simultaneous recordings during sound stimuli for representative azimuths. The simultaneously imaged units are sorted in the correlation matrices based on cross validated hierarchical clustering (see methods). Left bottom: Distribution histograms for the pairwise correlation coefficients (Kendall tau) from pairs of simultaneously recorded top ranked units across mice (blue) compared to the chance level distribution obtained through randomization of the temporal structure of each unit's activity to break correlations (purple). Vertical dashed lines show the medians of these distributions. *: $p < 0.05$, ***: $p < 0.0001$, Kolmogorov-Smirnov with Sidak. Right: Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained using a Bayes classifier (naive approximation for computation efficiency) to decode the single-trial response patterns from the 6 (neuropixels) or 7 (sTeFo 2P imaging) top ranked units in the simultaneously acquired datasets across mice (cyan), modeled decorrelated datasets (orange) and the chance level distribution associated with our stimulation paradigm (gray).

the CNIC of passively listening animals (Day et al., 2012 [↗](#); Day and Delgutte, 2016 [↗](#), 2013 [↗](#)), which could have potentially exciting implications, as DCIC represents a higher order relay of the auditory pathway, with respect to CNIC, and receives more profuse descending corticofugal inputs involved in sound location experience dependent plasticity (Bajo et al., 2019 [↗](#), 2010 [↗](#); Lesicko et al., 2022 [↗](#); Winer et al., 2002 [↗](#)). Concretely, we show that sound location coding does indeed occur at DCIC on the single trial basis, and that this follows a comparable mechanism to the characterized population code at CNIC (Day and Delgutte, 2013 [↗](#)). However, it remains to be determined if indeed the DCIC network is physiologically capable of Bayesian decoding computations. Interestingly, the small number of DCIC top ranked units necessary to effectively decode stimulus azimuth suggests that sound azimuth information is redundantly distributed across DCIC top ranked units, which points out that mechanisms beyond coding efficiency could be relevant for this population code.

While the decoding error observed from our DCIC datasets obtained in passively listening, untrained mice coincidentally matches the discrimination ability of highly trained, motivated mice (Lauer et al., 2011 [↗](#)), a relationship between decoding error and psychophysical performance remains to be determined. Interestingly, a primary sensory representations should theoretically be even more precise than the behavioral performance as reported in the visual system (Stringer et al., 2021 [↗](#)). One possible explanation could be that the population code of sound azimuth at DCIC is likely not a primary representation, as DCIC is reported to be involved in higher order functions including experience dependent plasticity, is influenced by auditory cortex (Bajo et al., 2019 [↗](#), 2010 [↗](#); Bajo and King, 2012 [↗](#)) and is associated to non-auditory processes (Gruters and Groh, 2012 [↗](#); Wong and Borst, 2019 [↗](#)). In this respect, we observed a correlation between the recorded DCIC population activity and snout movements, suggesting that non-auditory processes are likely influencing the recorded responses (Suppl. Fig. 2 [↗](#)). On the other hand, our observations were drawn from unilateral datasets from a single IC. It has been reported that the population responses of a single IC (unilateral recordings from CNIC) can produce a complete representation of sound location throughout the frontal hemifield (Day and Delgutte, 2016 [↗](#), 2013 [↗](#)). Our observations support this notion, showing that unilateral subpopulations of top ranked DCIC units carry enough information to generate a representation of the complete frontal hemifield that is accurate enough to match the discrimination ability of mice (Lauer et al., 2011 [↗](#)). Discrepancies with respect to unilateral lesion studies (Day and Delgutte, 2013 [↗](#); Jenkins and Masterton, 1982 [↗](#)), suggesting that one IC would carry information about the contralateral hemifield only, could be explained by the loss of commissural interactions between both ICs that contribute to sound location processing (Orton et al., 2016 [↗](#)). In this respect, we observed that the jRGECO1a-labeled neurons we imaged could project contralaterally (Fig. 1A [↗](#), inset), suggesting that the imaged DCIC populations would be involved in commissural interactions. This is particularly interesting in the context of the opposing two channel model of sound location coding proposed for the mammalian auditory brainstem (Grothe et al., 2010 [↗](#); Park et al., 2004 [↗](#)), which was ruled out for azimuth coding at the rabbit CNIC (Day and Delgutte, 2013 [↗](#)), but could still be relevant for understanding DCIC population coding of sound location. Thus, we cannot rule out that the actual accuracy of the DCIC population code for sound azimuth, on the single-trial basis, would only be reached if we considered bilateral DCIC recordings. These open questions about the transformations undergone by the neural code for sound location across the relays of the auditory pathway and the influence of bilateral interactions are exciting subjects for future study.

Both our imaging and neuropixels datasets show that the responses from DCIC neurons are noisy, independently of the recording methodology or stimulus (Fig. 4 [↗](#), 5, Suppl. Fig. 1 [↗](#)), and that this neuronal noise is often positively correlated across simultaneously recorded pairs of top ranked units, which are involved in the population code of sound azimuth (Fig. 6 [↗](#)). Interestingly, these noise correlations can contribute to reducing the error of the DCIC population code for sound azimuth (Fig. 6 [↗](#)). In low dimensional samples of ~two neurons this could only be justified by concomitant negative signal correlations with positively correlated noise (middle scenario in figure 5A [↗](#)). However, the recorded top ranked units typically showed contralateral sensitivity,

which means generally positive signal correlation instead. Nevertheless, in larger samples this relationship between signal and noise correlations seems to be more complex, supporting that these assumptions do not necessarily apply in a complex multidimensional representation like the population code here studied, where empirical determination of the impact of noise correlations are more informative (Montijn et al., 2016). Noise correlations have been reported to occur at the mammalian CNIC and have a role in sound categorization (Sadeghi et al., 2019), however their influence on the population code for sound location remained unexplored (Day et al., 2012; Day and Delgutte, 2016, 2013). Thus, our data supports that response sensitivity to azimuth might not be the only feature of IC neurons carrying information about stimulus azimuth, as considered in previous studies (Chase and Young, 2008, 2005; Day et al., 2012; Day and Delgutte, 2016, 2013), but noise correlations across the population responses could also be a relevant factor. This implication might extend to other brain regions relevant for sound location coding or perhaps also to different sensory modalities.

It is worth mentioning that many discrepancies exist in the labeling and segmentation boundaries of the anatomical subdivisions of IC between the most commonly used mouse brain reference atlases (Chon et al., 2019), like the Franklin-Paxinos atlas (Paxinos and Franklin, 2001) and the Allen atlas (atlas.brain-map.org), which can lead to conflicting interpretations of experimental data (Bjerke et al., 2018). These differences could arise due to different criteria employed by the expert neuroanatomists to segment anatomical subdivisions (histological stainings or magnetic resonance imaging) different tissue preparation (PFA fixed vs. fresh frozen) or intrinsic variability in the colonies of the animals employed (Chon et al., 2019). Here we adopted the IC segmentation boundaries and labels from the Franklin-Paxinos atlas, due to the extensive previous research employing it and the fact that we followed a compatible tissue preparation procedure for determining the anatomical location of our imaged volumes and electrode tracts (PFA fixed tissue slices). Recent efforts implementing the widely used segmentation and labeling from the Franklin-Paxinos atlas into the Allen common coordinate framework are a step in the right direction to circumvent this issue (Chon et al., 2019). Taking this into consideration, we report that DCIC top ranked units scatter across the imaged volume with no specificity in their location across the anatomical axes (Fig. 6). Interestingly, *in vivo* imaging studies report that the mouse dorsal IC processes sound frequency information topographically (tonotopy), showing a distinct tonotopic arrangement of neurons across the rostromedial-caudolateral and dorso-ventral axes (Barnstedt et al., 2015; Wong and Borst, 2019). Beyond the mouse model, a hypothetical functional relationship between the distribution of sound location encoding DCIC neuronal subpopulations and DCIC tonotopic gradients of sound frequency tuning could shed light into how the DCIC subpopulation code of sound azimuth relates to the duplex theory of sound location (Rayleigh, 1907). Nevertheless, we observed that not all DCIC azimuth sensitive units show pure tone sound frequency sensitivity, whereas a fraction of units showing both azimuth and sound frequency sensitivity displayed selectivity for high sound frequencies > 10kHz, making this argument only relevant to perhaps the latter subset of DCIC neurons. Technical improvements realizing larger fields of view to interrogate the wide-stretching tonotopic DCIC map in simultaneous recordings would be key to further explore this relationship (Demas et al., 2021; Ota et al., 2021).

In conclusion, our simultaneous recordings from passively listening mice directly support that population response patterns from a subset of DCIC noisy neurons effectively encode sound location on the single-trial basis with an error that matches the discrimination ability of trained mice (Lauer et al., 2011). Specifically, we report that azimuth information is redundantly distributed across the subset of DCIC azimuth sensitive units (top ranked units), which rely not only on response dependency to sound location but also on noise-correlations to accurately encode this information and are randomly scattered across DCIC. An important open question remains about the behavioral relevance of this code and the noise correlations. We hope that this

study paves the way to explore this question in detail, together with the wealth of knowledge available about cortico-collicular interactions involved in experience dependent plasticity in the context of sound localization (Bajo et al., 2019 [DOI](#), 2010 [DOI](#)).

Materials and methods

Animals and ethics statement

This work was performed in compliance to the European Communities Council Directive (2010/63/EU) to minimize animal pain and discomfort. EMBL's committee for animal welfare and institutional animal care and use (IACUC) approved all experimental procedures under protocol number 2019-04-15RP. Experiments were performed on 7-16-week-old CBA/j mice obtained from Charles River Laboratories and housed in groups of 1-5 in makrolon type 2L cages on ventilated racks at room temperature and 50% humidity with a 12 hr light cycle. Food and water were available ad libitum. Experimental subjects consisted of 12 mice for imaging and 4 mice for electrophysiological recordings.

Surgical procedures

For microscopy

Cranial window surgeries were performed on 7-8-week-old mice of either sex following procedures published elsewhere (Boffi et al., 2018 [DOI](#)). Briefly, anesthesia consisted of a mixture of 40 μ l fentanyl (0.1 mg/ml; Janssen), 160 μ l midazolam (5 mg/ml; Hameln) and 60 μ l medetomidin (1 mg/ml; Pfizer), dosed in 5 μ l/g body weight and injected i.p.. Hair removal cream was used to remove the fur over the scalp of anesthetized mice and eye ointment was applied (Bepanthen, Bayer). 1% xylocain (AstraZeneca) was injected as preincisional anesthesia under the scalp. Prepared mice were then placed in a stereotaxic apparatus (David Kopf Instruments, model 963) with their bodies on a heating pad (37°C). The scalp was surgically removed to expose the dorsal cranium. The periosteum was removed with fine forceps and scissors to prepare the surface for cranial window implantation and stereotaxic delivery of viral vectors. Post-surgical pain relief was provided (Metacam, Boehringer Ingelheim) through s.c. injections (0.1 mg/ml, dosed 10 μ l/g body weight).

The Ca^{2+} indicator jRGECO1a (Dana et al., 2016 [DOI](#)) was expressed in IC neurons through transduction with AAV vectors (Addgene #100854-AAV1), which were stereotaxically delivered as follows. A 4mm diameter circular craniectomy centered over IC (~1mm posterior to lambda (34)) was produced using a dental drill (Microtorque, Harvard Apparatus) with care to avoid bleeding and damage of the dura, which was not removed and left intact. Stereotaxic injections were performed at the right IC (from bregma: -5.2 mm AP, 0.5 mm ML) with pulled glass pipettes lowered to depths of 300, 400 and 500 μ m, at a rate of ~4 μ l/hr using a 10 ml syringe (to generate pressure) coupled to the glass needle through a silicon tubing via a luer 3-way T valve. ~300nl were injected per site. After injection, the craniectomy was sealed with a round 4mm coverslip (~170 μ m thick, disinfected with 70% ethanol), with a drop of saline between the glass and the dura, and dental acrylic cement (Hager Werken Cyano Fast and Paladur acrylic powder). A head fixation bar was also cemented. The open skin wound was also closed with acrylic cement. At the end of the surgery anesthesia was antagonized with a subcutaneous injection of a mixture of 120 μ l sterile saline, 800 μ l flumazenil (0.1 mg/ml; Fresenius Kabi), and 60 μ l atipamezole (5 mg/ml; Pfizer) dosed in 10 μ l/g body weight. Mice were single housed after surgery to minimize the chance of cage mates compromising each others implantations and were allowed to recover for at least 4 weeks before imaging, providing time for Ca^{2+} indicator expression and for the inflammation associated with this surgery to resolve (Holtmaat et al., 2009 [DOI](#)).

For electrophysiology

Acute craniectomy surgeries were performed on 7-15-week-old mice of either sex following procedures published elsewhere (Boffi et al., 2018 [DOI](#)). Briefly, anesthesia was induced with 5% isoflurane (Baxter) in O₂, and maintained at 1.5-2% with a flow rate of 0.4-0.6 LPM. Hair removal cream was used to remove the fur over the scalp, eye ointment was applied (Bepanthen, Bayer) and 1% xylocain (AstraZeneca) was injected under the scalp as preincisional anesthesia. Mice were set in the stereotaxic apparatus with their bodies on a heating pad (37°C) and their scalp and periosteum removed to expose the dorsal cranium. A custom made head bar was cemented to the exposed cranium, sealing the skin wound, with UV cured dental acrylic to reduce curing time (Loctite 4305 LC). A ~4mm diameter circular craniectomy centered over IC (~1mm posterior to lambda (Paxinos and Franklin, 2001 [DOI](#))) was produced using a dental drill (Microtorque, Harvard Apparatus) with care to avoid bleeding and damage of the dura. The surface of the brain was kept moist at all times with sterile cortex buffer (mM: NaCl 125, KCl 5, Glucose 10, Hepes 10, CaCl₂ 2, MgCl₂ 2, pH 7.4). The dura was carefully ripped open at a small site over DCIC for electrode insertion, through an incision made with a sterile G27 needle. The dura was not completely removed and the rest was left intact. Post-surgical pain relief was provided (Metacam, Boehringer Ingelheim) through s.c. injections (0.1 mg/ml, dosed 10µl/g body weight).

Craniectomy surgeries typically lasted 30 min since the induction of anesthesia, after which the mouse was allowed to recover headfixed at the recording setup (same as for Ca²⁺ imaging, see below). Typically ~5 minutes after isoflurane removal the mice displayed awake behaviors like whisking, blinking, grooming and body movements. The craniectomy was kept submerged in a well of cortex buffer throughout the experiment. We began recording at least 40 minutes after the mouse showed signs of being fully awake. During this period the neuropixel probe was inserted into the DCIC at 10µm/s rate using a micromanipulator (Sutter MPC-385 system, coupled to a Sensapex uMp-NPH Neuropixel 1.0 probe holder). After insertion, neuropixel probes were allowed to settle in the tissue for 10 min. The metal headbar was left exposed to the cortex buffer close to the edge of the craniectomy and was used as reference.

sTeFo 2P microscopy

We built a bespoke scanned temporal focusing (sTeFo) 2 photon microscope for fast volumetric in vivo Ca²⁺ imaging of large IC neuronal populations from awake mice. Technical details and detailed working principle are extensively described elsewhere (Prevedel et al., 2016 [DOI](#)). The main difference with respect to the microscope design published by Prevedel et al. (2016) [DOI](#) was a higher repetition laser (10 MHz, FemtoTrain, Spectra Physics). Laser power while scanning was kept below 191 mW, measured after the objective, to avoid heating the brain tissue excessively (Prevedel et al., 2016 [DOI](#)). 470 µm² fields of view were imaged at 128 px² resolution and spaced in 15 µm z steps to cover 570 µm in depth (38 z steps) at a volume rate of 4.01 Hz. Artifacts produced by objective piezo z drive flyback were excluded from analysis.

In a typical imaging session, cranial window implanted mice were briefly (< 1 min) anesthetized with 5% isoflurane in O₂ for quick head fixation at a custom stage, slightly restraining their bodies inside a 5 cm diameter acrylic tube, and positioned under our custom 2P microscope. Mice fully recovered from the brief isoflurane anesthesia, showing a clear blinking reflex, whisking and sniffing behaviors and normal body posture and movements, immediately after head fixation. A pilot group of mice were habituated to the head fixed condition in daily 20 min sessions for 3 days, however we did not observe a marked contrast in the behavior of habituated versus unhabituated mice during our relatively short 25 min imaging sessions. In consequence imaging sessions never surpassed a maximum of 25 min, after which the mouse was returned to its home cage. Typically, mice were imaged a total of 2-11 times (sessions), one to three times a week. Datasets here analyzed and reported come from the imaging session in which we observed maximal calcium sensor signal (peak AAV expression) and maximum number of detected units. Volumetric imaging

data was visualized and rendered using FIJI (Schindelin et al., 2012 [DOI](#)). Motion correction was performed using NoRMCorre (Pnevmatikakis and Giovannucci, 2017 [DOI](#)) on individual imaging planes from the volumetric datasets. We used the CaImAn package (Giovannucci et al., 2019 [DOI](#)) for automatic ROI segmentation through constrained non negative matrix factorization and selected ROIs (Units) showing clear Ca transients consistent with neuronal activity, and IC neuron somatic shape and size (Schofield and Beebe, 2019 [DOI](#)). jRGECO1a signal from the segmented ROIs was extracted, denoised and neuropil corrected from each individual imaging plane using CaImAn (Giovannucci et al., 2019 [DOI](#)). Spike probabilities were estimated using MLSpike (Deneux et al., 2016 [DOI](#)).

Extracellular electrophysiological recordings

We performed acute extracellular single-unit recordings using the Neuropixels (Jun et al., 2017 [DOI](#)) 1.0 system (PRB_1_4_0480_1, PXIe_1000, HST_1000, IMEC). Reference and ground pads on the band connector of the probe were bridged and grounded. The probe shank was coated before recordings by dipping 10 times every 5 s in DiI (V22885, ThermoFisher) for post hoc histological assessment of electrode tracks. Recordings were performed using SpikeGLX software (<https://billkarsh.github.io/SpikeGLX/> [DOI](#)). Mice were recorded for ~25min after successful electrode insertion (see surgical procedures), after which they were placed in an empty cage, and their behavior monitored for 10-15 minutes. We only considered recordings from mice that displayed normal behaviors (locomotion, exploration, sniffing, whisking), with no clear signs of pain or discomfort (freezing, shivering, raised fur) at this stage. Mice were sacrificed after this with CO₂ for transcardial perfusion and tissue preparation to trace electrode tracks (see below). After retrieval, neuropixels probes were cleaned through immersion in 1% Tergazime (Z742918, Merck) for some minutes, followed by washes in distilled water and reused multiple times. Recordings were preprocessed for spike sorting with CatGT (<https://billkarsh.github.io/SpikeGLX/#catgt> [DOI](#)) for high pass filtering, common average referencing of the demultiplexed channels and file concatenation. Spike sorting with drift correction was performed using kilosort 2.5 (<https://github.com/MouseLand/Kilosort/releases/tag/v2.5> [DOI](#)). The output of kilosort was manually curated using Phy (<https://github.com/cortex-lab/phy> [DOI](#)), selecting only single-units with somatic AP waveforms with good waveform consistency (amplitude) across spikes, showing few or no refractory period violations (2ms refractory period violation time window) in their correlogram plots and firing a minimum of 150 times in 20 min of recording. Stimulus times collected as digital channel inputs and recorded clock signals from the neuropixels headstage were temporally aligned to a reference clock with Tprime (<https://billkarsh.github.io/SpikeGLX/#tprime> [DOI](#)).

Sound Stimulation

A custom sound-attenuating chamber was incorporated to the microscope/electrophysiology setup, allowing only the objective or cables to access the inside, to attenuate room and microscope scanner noise reaching the mouse. All possible sound reflecting surfaces were lined with 1 cm or 0.5 cm thick foam to preserve free-field conditions. SPL was measured using a free-field prepolarized measurement microphone (PCB Piezotronics, 378C01, with signal conditioner 482A21) placed on the setup stage in the position of the head of the mouse. For imaging experiments, background R.M.S. SPL measurements during imaging without sound stimuli delivery was (20-40 kHz) 44.83 dB (79.5 dB total SPL summed across the band). Sound stimulus (20-40 kHz) R.M.S. SPL during imaging was 56.83 dB (96.2 dB total SPL summed across the band), ensuring that SPL was at least 10 dB above background SPL for sound stimuli to be salient. Background R.M.S. SPL during imaging across the mouse hearing range (2.5-80kHz) was 44.53 dB.

Electrophysiological experiments were performed at the same setup, without the microscope running. For broadband stimuli applied in our electrophysiological experiments, background R.M.S. SPL (15-50 kHz) was 35.96 dB (89.9 dB total SPL summed across the band). Broadband stimulus (15-50kHz) R.M.S. SPL was 62.5 dB (102 dB total SPL summed across the band). Pure tone

stimuli consisted on 4-48kHz tones, making 3.5 octaves split into steps of 4. Peak SPL of the pure tone stimuli was calibrated to 65 ± 3 dB. R.M.S. background noise across the 4-48kHz band was 35.1 dB. Background R.M.S. SPL across the mouse hearing range (2.5-80kHz) was 34.68 dB.

Broadband sound stimuli consisted of band-passed frozen noise. Stimulus duration was 500 ms for imaging experiments and 200 ms for electrophysiological experiments, including 10 ms up and down ramps. Sound stimuli were delivered through an electrostatic speaker (Tucker Davis Technologies, ES1 coupled to a ED1 speaker driver) mounted on an arm at an 8.5 cm distance away from the head of the mouse, which was mounted on a NEMA17 bipolar stepper motor (26 Ncm, 1.8 deg/step) placed under the head of the mouse to position the speaker around the frontal hemifield. The motor was controlled using an Arduino UNO microcontroller, running GRBL v0.9 (G-code interpreter, <https://github.com/grbl/grbl>) and coupled with the Arduino CNC Shield V3.10 (<https://blog.protoneer.co.nz/arduino-cnc-controller/>) running A4988 motor drivers. To control and program speaker positioning, we used the universal G-code sender (https://winder.github.io/ugs_website/). Synchronization of speaker positioning and sound stimulus delivery to data acquisition was done through a second Arduino UNO board receiving a scan start TTL from the microscope to trigger stimulation programs or by recording the sound delivery TTL signal with a DAQ board (National Instruments, PXIe-6341) through a digital channel in SpikeGLX. Pure tone stimuli were presented with the speaker parked at 0° azimuth, in front of the mouse.

Audio stimuli (including 10 ms up and down ramps) were synthesized using MATLAB (0.5 MHz sampling rate). Sound delivery was performed using a DAQ board (National Instruments, PXIe-6341) interfaced to MATLAB using the DAQmx wrapper from scanimage (<https://vidriotechnologies.com/scanimage/>).

A typical stimulation protocol consisted of a series of presentations of a sound stimulus in which azimuth angles varied randomly from 13 different frontal positions (frontal 180° split into 15° steps). For imaging experiments, stimuli were presented every 5 seconds and each azimuth angle was presented on average 14 times per session. Minimum number of same-azimuth trials collected was 8. For electrophysiological experiments, firstly pure tone stimuli were applied every 2 seconds in pseudo-random order to obtain 10 trial repetitions (14 sound frequencies, 140 trials), and after that broadband stimulus azimuth trials were applied every 3 seconds (to allow the motor to complete speaker movements) in pseudo-random order to obtain 20 trial repetitions (13 azimuths, 260 trials).

Histology

At the end of the experiment mice were transcardially perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde at room temperature to fixate the brain. The perfused brain was dissected and post-fixed in 4% paraformaldehyde for 24 h at 4 °C. After post-fixation, brains were washed with PBS saline and stored at 4 °C. 100-μm-thick free-floating vibratome (Leica VT1200) coronal sections were cut at room temperature, mounted on slides using Vectashield with DAPI (H-1200-10, Vector Laboratories) and imaged in a NIKON Ti-E epifluorescence microscope equipped with a LUMENCOR SPECTRA X module for illumination, standard DAPI and mCherry filter cubes, and a pco.edge 4.2 CL sCMOS camera; using a CFI P-Apo 4x Lambda/ 0.20/ 20,00 objective. Images of brain slices were aligned to the mouse brain reference atlas manually following anatomical landmarks.

Face movement videography

Videos from one side of the mouse face, ipsilateral to the imaged IC, were recorded with an IR sensitive camera equipped with a CMOS OV2710 sensor, IR LEDs for illumination, a 25mm M12 objective (ELP, USBFHD05MT-KL36IR), using a sampling rate of 30 fps at 720p resolution. Video acquisition was synchronized to the microscope acquisition start. Video image analysis based on intensity change between successive frames was performed following a similar approach as

described by Wong and Borst (2019) [using FIJI](#) (Schindelin et al., 2012 [using](#)) and custom MATLAB scripts. For correlations between face movements and imaged neuronal activity, the Pearson correlation coefficient was used.

Statistical analyses

Each mouse was considered a replicate. Datasets obtained in one imaging/recording session from each mouse, consisting of ~200 azimuth trials were analyzed. We did not perform any form of data augmentation through resampling with restitution of the recorded responses to virtually increase the number of trials (Day and Delgutte, 2013 [using](#)). Statistical analyses were performed using MATLAB (2023a,b) functions and custom scripts. For hypothesis testing, a p value < 0.05 was considered significant. Sidak's correction of p values was used for multiple comparisons. nS/N ratios were evaluated as the inverse of the coefficient of variation of the recorded responses of each cell across azimuth trials.

For our imaging data, since many units failed to respond across many trials, we did not perform baseline spike probability subtraction as this frequently led to negative spike probabilities that were hard to interpret. For our neuropixels experiments, we generally did not perform baseline firing subtraction from the responses as the average baseline firing in the 200ms before stimuli was quite low and close to 0 (Suppl. Fig. 1F [using](#), Fig. 5A [using](#)). Nevertheless, we did subtract baseline firing for the determination of the nS/N (Fig. 5C [using](#)), as the on going activity varied extensively across the recordings (in the absence of sound stimulation, Fig. 1I [using](#)) and this could lead to nS/N values that were artificially high. On going activity was estimated as the spike count determined in 200ms time windows during the inter trial periods (2.8s, no sound stimulation), 1s before each stimulation trial. "Baseline on-going firing" was determined as the spike count 200ms previous to that. For fairness of comparisons, we also subtracted baseline firing determined during the 200ms pre response for nS/N determination from the sound stimulated condition.

One sided Kruskal-Wallis tests were performed to determine median response tuning to stimulus azimuth, with 12 degrees of freedom (13 tested azimuths). Two sample Kolmogorov-Smirnov and Wilcoxon rank sum tests were two sided. To find units with single-trial responses that were statistically dependent on stimulus azimuth (feature selection) we performed χ^2 independency tests using the function `fschi2` from the statistics and machine learning toolbox from MATLAB with default parameters, which binned the responses into 10 bins. Cross validated sorting of noise correlation matrices was obtained by performing hierarchical clustering on half of the trials recorded based on pairwise neuronal noise correlations (using the Kendall tau as a non parametric correlation coefficient) as a distance metric, and plotting the pairwise correlation coefficient matrix registered for the other half of the trials, sorting the units based on the clustering. IC population response classification (decoding) was performed using the naive Bayes classifier algorithm implemented in MATLAB, using the `fitcnb` function from the statistics and machine learning toolbox. For the latter function, model fit involved Bayesian hyperparameter optimization for all eligible parameters. This MATLAB function performs 3 steps: 1) Estimation of the densities of the units' responses (predictors) within each class (stimulus azimuth) through kernel smoothing density estimation. Uniform, Epanechnikov, Gaussian or triangular smoothing kernels and Kernel smoothing window width were set through hyperparameter optimization. 2) Use the Bayes rule to estimate the posterior probability $\hat{P}$ for all possible azimuth classes $k=1, \dots, 13$ (-90° to 90° in 15° steps) as:

$$\hat{P}(Y=k|X_1 \dots X_p) = \frac{\pi(Y=k) \prod_{j=1}^p P(X_j|Y=k)}{\sum_{k=1}^{13} \pi(Y=k) \prod_{j=1}^p P(X_j|Y=k)},$$

where Y is the possible azimuth class being evaluated, $X_1, \dots, X_p$ are the simultaneously acquired responses from the units in a population of size p during a single azimuth trial, $\pi(Y=k)$ is the prior probability that the azimuth class is k , determined as the relative frequencies of each azimuth class in the training dataset. 3) Determine the azimuth class of a single trial response pattern as the class with maximum posterior probability (maximum *a posteriori* decision rule).

For dimensionality reduction, feature selection and classification (decoding) a “leave one out” cross-validation strategy was implemented for the whole process, in which dimensionality reduction (PCA) or feature selection (top ranked cell selection, if applicable) was performed and then the models were fitted to the population responses collected in all trials except one (for each imaging session) and the one population response left out was decoded as is or using the selected features (best cells) or using the PCA loadings matrix to calculate PC scores, and fitted model. This procedure was iterated until all population responses simultaneously recorded in an experiment (mouse) were decoded.

Acknowledgements

We acknowledge Lina Streich and Ling Wang for their contribution to microscope assembly and setup, the mechanical and electronics workshops and the laboratory animal resource facility of EMBL Heidelberg for technical assistance. Evan Harrell for input on motorized speaker design and Jacques Bourg for input on IC cranial window surgeries. Peter Rupprecht and Alejandro Tlaie-Boria for helpful comments on the manuscript. JCB acknowledges supporting fellowships from the EMBL Interdisciplinary Postdoc (EIPOD) Programme under Marie Skłodowska Curie Cofund Actions MSCA-COFUND-FP (664726). This work was funded by the European Molecular Biology Laboratory (EMBL) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project 458898724 awarded to RP.

Additional information

Authorship contributions

JCB, BB, HA, and RP conceptualized the study. JCB performed experiments. JCB, BB, HA and RP analyzed and interpreted data. JCB wrote preliminary manuscript draft and prepared figures. JCB, BB, HA, and RP edited and finalized manuscript.

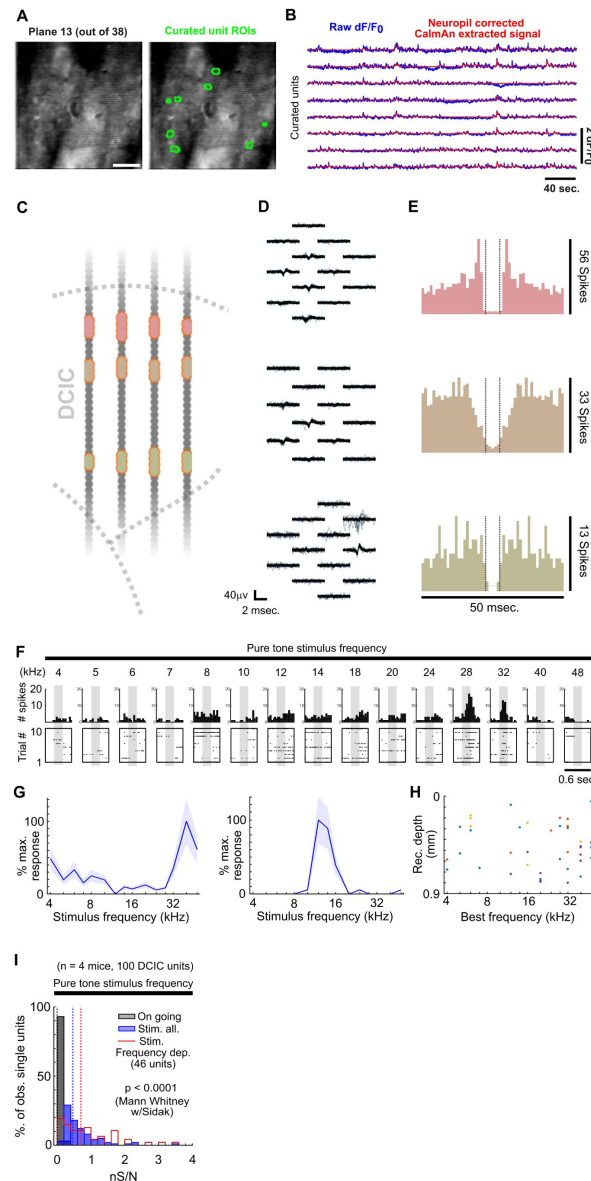
Competing interests

The authors declare no competing financial interests.

Materials & correspondence

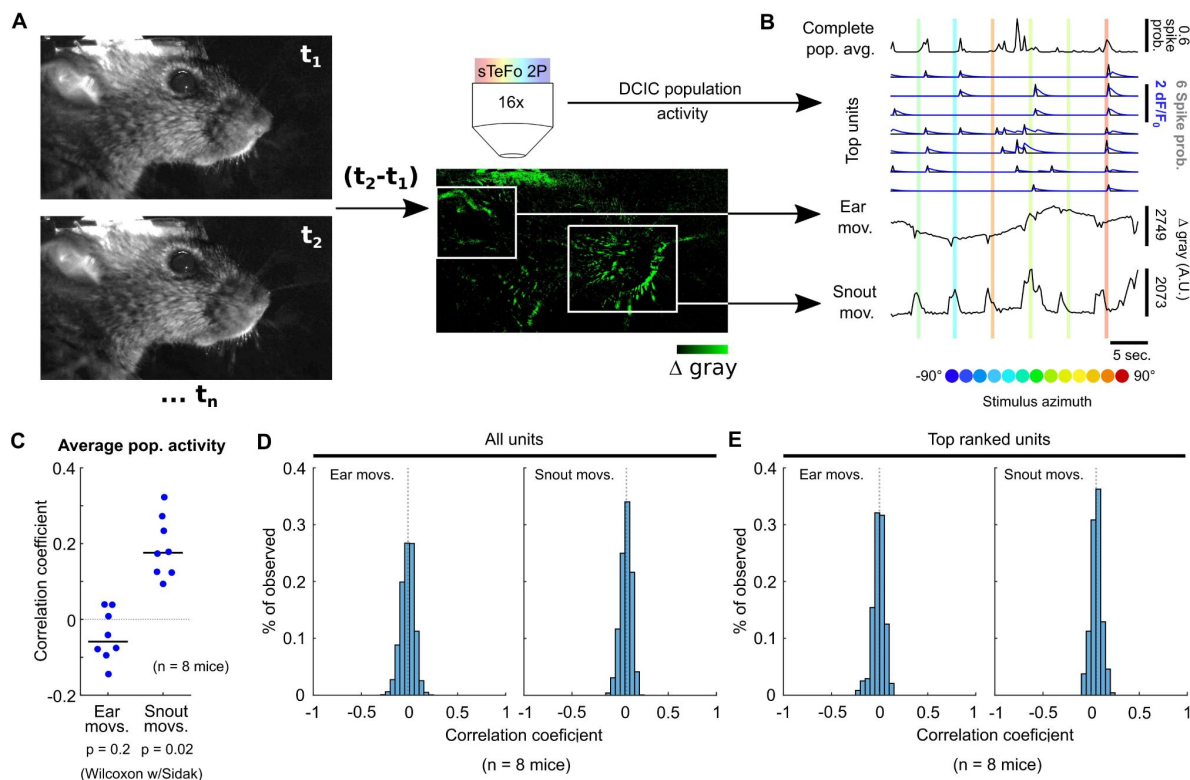
Datasets and code supporting this study are available from <https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1064> .

Supplemental figures



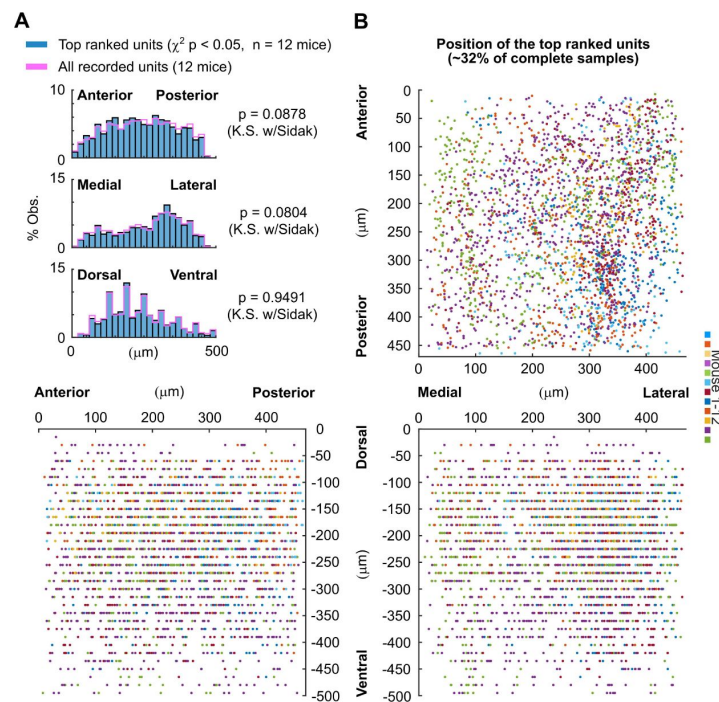
Supplemental figure 1

A) Representative median intensity projection from one sTeFo 2P imaging plane timelapse, extracted from one of our volumetric imaging datasets (left panel, scale bar 100 μ m) with superimposed footprints from CaImAn segmented and manually curated ROIs corresponding to recorded units (right panel). B) Representative raw dF/F_0 traces extracted from CaImAn segmented and manually curated ROIs (units, blue) with CaImAn neuropil corrected and denoised output superimposed (red). C) Schematic representation of a neuropixels probe shank location across DCIC (approximated histologically) highlighting in different colors the channels across which curated single-units were detected after spike sorting with kilosort 2.5 and manual curation with Phy. D) 100 superimposed traces from aligned spikes recorded from the color coded channels highlighted in (C), displaying spike amplitude and somatic waveform consistency and spatial attenuation. E) Correlogram plots from the curated single-units shown in (C,D), displaying very few refractory period violations (vertical dotted lines: 2ms refractory period time window considered). F) Representative responses to sound frequency (pure tone stimuli) of a frequency sensitive single-unit. Top panels show the peri-stimulus time histograms and the bottom panels show the corresponding spike raster plots across trials. G) Representative sound frequency tuning curves from two DCIC sound frequency dependent single-units recorded with neuropixels. Mean and standard deviation are plotted. H) Relationship between the best frequency of DCIC sound frequency dependent single-units and recording depth. Data from n = 4 mice, point color corresponds to the same mouse. X axis is in logarithmic scale. I) Neuronal signal to noise level (nS/N) histograms from neuropixels recorded DCIC single-units in the absence of sound stimuli (on going activity, gray) and in response to pure tone stimuli trials at the best frequency for all collected single-units (blue) and the sound frequency sensitive subset (red).



Supplemental figure 2

A) Schematic representation of the strategy employed to videographically track face movements (ear and snout) during imaging experiments. B) Representative traces of imaged population activity and ear or snout movement across stimulus azimuth trials. C) Correlation between the average population activity trace and ear or snout movements across mice. Median value across mice is represented by a horizontal line. D, E) Histograms of the correlation coefficients between ear or snout movement and the imaged activity of each unit (D) or the top ranked unit subset (azimuth dependent, E) across mice.



Supplemental figure 3

Top ranked units are scattered throughout the imaged DCIC volumes.

A) Histogram plots of the distribution of top ranked unit position in the imaged volume across each anatomical axis (20 μm bins) obtained from all imaged mice, either for the complete sample of units (purple) or the subsample of top ranked units (~32% of all the units imaged per mice, blue). K.S.: Kolmogorov-Smirnov test. B) Scatter plots of the centroid position throughout the anatomical axes in the imaged volume from the detected top ranked units across mice.

References

- Adams JC (1979) **Ascending projections to the inferior colliculus** *J Comp Neurol* **183**:519–538 <https://doi.org/10.1002/cne.901830305>
- Averbeck BB, Latham PE, Pouget A (2006) **Neural correlations, population coding and computation** *Nat Rev Neurosci* **7**:358–366 <https://doi.org/10.1038/nrn1888>
- Bajo VM, King AJ (2012) **Cortical modulation of auditory processing in the midbrain** *Front Neural Circuits* **6**:1–12 <https://doi.org/10.3389/fncir.2012.00114>
- Bajo VM, Nodal FR, Korn C, Constantinescu AO, Mann EO, Boyden ES, King AJ (2019) **Silencing cortical activity during sound-localization training impairs auditory perceptual learning** *Nat Commun* **10** <https://doi.org/10.1038/s41467-019-10770-4>
- Bajo VM, Nodal FR, Moore DR, King AJ (2010) **The descending corticocollicular pathway mediates learning-induced auditory plasticity** *Nat Neurosci* **13**:253–260 <https://doi.org/10.1038/nn.2466>
- Barnstedt O, Keating P, Weissenberger Y, King AJ, Dahmen JC (2015) **Functional Microarchitecture of the Mouse Dorsal Inferior Colliculus Revealed through In Vivo Two-Photon Calcium Imaging** *J Neurosci* **35**:10927–10939 <https://doi.org/10.1523/JNEUROSCI.0103-15.2015>
- Bjerke IE, Øvsthus M, Andersson KA, Blixhavn CH, Kleven H, Yates SC, Puchades MA, Bjaalie JG, Leergaard TB (2018) **Navigating the murine brain: Toward best practices for determining and documenting neuroanatomical locations in experimental studies** *Front Neuroanat* **12**:1–15 <https://doi.org/10.3389/fnana.2018.00082>
- Boffi JC, Knabbe J, Kaiser M, Kuner T (2018) **KCC2-dependent Steady-state Intracellular Chloride Concentration and pH in Cortical Layer 2/3 Neurons of Anesthetized and Awake Mice** *Front Cell Neurosci* **12**:1–14 <https://doi.org/10.3389/fncel.2018.00007>
- Brunso-Bechtold JK, Thompson GC, Masterton RB (1981) **HRP study of the organization of auditory afferents ascending to central nucleus of inferior colliculus in cat** *J Comp Neurol* **197**:705–722 <https://doi.org/10.1002/cne.901970410>
- Chase SM, Young ED (2008) **Cues for Sound Localization Are Encoded in Multiple Aspects of Spike Trains in the Inferior Colliculus** *J Neurophysiol* **99**:1672–1682 <https://doi.org/10.1152/jn.00644.2007>
- Chase SM, Young ED (2005) **Limited segregation of different types of sound localization information among classes of units in the inferior colliculus** *J Neurosci* **25**:7575–7585 <https://doi.org/10.1523/JNEUROSCI.0915-05.2005>
- Chon U, Vanselow DJ, Cheng KC, Kim Y (2019) **Enhanced and unified anatomical labeling for a common mouse brain atlas** *Nat Commun* **10** <https://doi.org/10.1038/s41467-019-13057-w>

- Clarey JC, Barone P, Irons WA, Samson FK, Imig TJ (1995) **Comparison of Noise and Tone Azimuth Tuning of Neurons in Cat Primary Auditory-Cortex and Medial Geniculate-Body** *J Neurophysiol* **74**:961–980 <https://doi.org/10.1152/jn.1995.74.3.961>
- Clough M, Chen IA, Park SW, Ahrens AM, Stirman JN, Smith SL, Chen JL (2021) **Flexible simultaneous mesoscale two-photon imaging of neural activity at high speeds** *Nat Commun* **12**:1–7 <https://doi.org/10.1038/s41467-021-26737-3>
- Dana H *et al.* (2016) **Sensitive red protein calcium indicators for imaging neural activity** *Elife* **5**:1–24 <https://doi.org/10.7554/eLife.12727>
- Dana H, Shoham S (2012) **Numerical evaluation of temporal focusing characteristics in transparent and scattering media: erratum** *Opt Express* **20** <https://doi.org/10.1364/oe.20.028281>
- Day ML, Delgutte B (2016) **Neural population encoding and decoding of sound source location across sound level in the rabbit inferior colliculus** *J Neurophysiol* **115**:193–207 <https://doi.org/10.1152/jn.00643.2015>
- Day ML, Delgutte B (2013) **Decoding Sound Source Location and Separation Using Neural Population Activity Patterns** *J Neurosci* **33**:15837–15847 <https://doi.org/10.1523/JNEUROSCI.2034-13.2013>
- Day ML, Koka K, Delgutte B (2012) **Neural encoding of sound source location in the presence of a concurrent, spatially separated source** *J Neurophysiol* **108**:2612–2628 <https://doi.org/10.1152/jn.00303.2012>
- Demas J, Manley J, Tejera F, Barber K, Kim H, Traub FM, Chen B, Vaziri A (2021) **High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy** *Nature Methods* **18**:1103–1111 <https://doi.org/10.1038/s41592-021-01239-8>
- Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, Rozsa B, Vanzetta I (2016) **Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo** *Nat Commun* **7** <https://doi.org/10.1038/ncomms12190>
- Felix RA, Gourévitch B, Portfors C V (2018) **Subcortical pathways: Towards a better understanding of auditory disorders** *Hear Res* **362**:48–60 <https://doi.org/10.1016/j.heares.2018.01.008>
- Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) **ON THE RELATIONS BETWEEN THE DIRECTION TWO-DIMENSIONAL ARM MOVEMENTS AND CELL DISCHARGE PRIMATE MOTOR CORTEX1 IN** *J Neurosci* **2**:1527–1537
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) **Neuronal Population Coding of Movement Direction** *Science* **233**:1416–1419
- Giovannucci A *et al.* (2019) **CaImAn an open source tool for scalable calcium imaging data analysis** *Elife* **8**:1–45 <https://doi.org/10.7554/eLife.38173>
- Groh JM, Kelly KA, Underhill AM (2003) **A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus** *J Cogn Neurosci* **15**:1217–1231 <https://doi.org/10.1162/089892903322598166>

- Grothe B, Pecka M, McAlpine D (2010) **Mechanisms of Sound Localization in Mammals** *Physiol Rev* **90**:983–1012 <https://doi.org/10.1152/physrev.00026.2009>
- Gruters KG, Groh JM (2012) **Sounds and beyond: Multisensory and other non-auditory signals in the inferior colliculus** *Front Neural Circuits* **6**:1–39 <https://doi.org/10.3389/fncir.2012.00096>
- Guo F, Yu J, Jung HJ, Abruzzi KC, Luo W, Griffith LC, Rosbash M (2016) **Circadian neuron feedback controls the Drosophila sleep-activity profile** *Nature* **536**:292–297 <https://doi.org/10.1038/nature19097>
- Holtmaat A *et al.* (2009) **Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window** *Nat Protoc* **4**:1128–1144 <https://doi.org/10.1038/nprot.2009.89>
- Huang L *et al.* (2021) **Relationship between simultaneously recorded spiking activity and fluorescence signal in gcamp6 transgenic mice** *Elife* **10**:1–19 <https://doi.org/10.7554/eLife.51675>
- Jazayeri M, Movshon JA (2006) **Optimal representation of sensory information by neural populations** *Nat Neurosci* **9**:690–696 <https://doi.org/10.1038/nn1691>
- Jenkins WM, Masterton B (1982) **Sound Localization: Effects Of Unilateral Lesions in Central Auditory System** *J Neurophysiol* **47**
- Jun JJ *et al.* (2017) **Fully integrated silicon probes for high-density recording of neural activity** *Nature* **551**:232–236 <https://doi.org/10.1038/nature24636>
- King AJ, Schnupp JWH, Doubell TP (2001) **The shape of ears to come: Dynamic coding of auditory space** *Trends Cogn Sci* **5**:261–270 [https://doi.org/10.1016/S1364-6613\(00\)01660-0](https://doi.org/10.1016/S1364-6613(00)01660-0)
- Knudsen EI, Konishi M (1979) **Mechanisms of sound localization in the barn owl (Tyto alba)** *J Comp Physiol* **133**:13–21 <https://doi.org/10.1007/BF00663106>
- Kohn A, Coen-Cagli R, Kanitscheider I, Pouget A (2016) **Correlations and Neuronal Population Information** *Annu Rev Neurosci* **39**:237–256 <https://doi.org/10.1146/annurev-neuro-070815-013851>
- Kulkarni A, Colburn HS (1998) **Role of spectral detail in sound-source localization** *Nature* **396**:747–749 <https://doi.org/10.1038/25526>
- Lauer AM, Slee SJ, May BJ (2011) **Acoustic basis of directional acuity in laboratory mice** *JARO - J Assoc Res Otolaryngol* **12**:633–645 <https://doi.org/10.1007/s10162-011-0279-y>
- Lesica NA, Lingner A, Grothe B (2010) **Population coding of interaural time differences in gerbils and barn owls** *J Neurosci* **30**:11696–11702 <https://doi.org/10.1523/JNEUROSCI.0846-10.2010>
- Lesicko AMH, Angeloni CF, Blackwell JM, De Biasi M, Geffen MN (2022) **Cortico-Fugal Regulation of Predictive Coding** *Elife* **11**:1–29 <https://doi.org/10.7554/eLife.73289>
- Montijn JS, Meijer GT, Lansink CS, Pennartz CMA (2016) **Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective** *Cell Rep* **16**:2486–2498 <https://doi.org/10.1016/j.celrep.2016.07.065>

- Orton LD, Papasavvas CA, Rees A (2016) **Commissural Gain Control Enhances the Midbrain Representation of Sound Location** *J Neurosci* **36**:4470–4481 <https://doi.org/10.1523/JNEUROSCI.3012-15.2016>
- Ota K *et al.* (2021) **Fast, cell-resolution, contiguous-wide two-photon imaging to reveal functional network architectures across multi-modal cortical areas** *Neuron* **109**:1810–1824 <https://doi.org/10.1016/j.neuron.2021.03.032>
- Pachitariu M, Stringer C, Harris KD (2018) **Robustness of spike deconvolution for neuronal calcium imaging** *J Neurosci* **38**:7976–7985 <https://doi.org/10.1523/JNEUROSCI.3339-17.2018>
- Panniello M, King AJ, Dahmen JC, Walker KMM (2018) **Local and global spatial organization of interaural level difference and frequency preferences in auditory cortex** *Cereb Cortex* **28**:350–369 <https://doi.org/10.1093/cercor/bhx295>
- Park TJ, Klug A, Holinstat M, Grothe B (2004) **Interaural level difference processing in the lateral superior olive and the inferior colliculus** *J Neurophysiol* **92**:289–301 <https://doi.org/10.1152/jn.00961.2003>
- Paxinos G, Franklin K (2001) **The Mouse brain in Stereotaxic Coordinates, Second Edi.** ed Academic Press
- Pnevmatikakis EA, Giovannucci A (2017) **NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data** *J Neurosci Methods* **291**:83–94 <https://doi.org/10.1016/j.jneumeth.2017.07.031>
- Prevedel R *et al.* (2016) **Fast volumetric calcium imaging across multiple cortical layers using sculpted light** *Nat Methods* **13**:1021–1028 <https://doi.org/10.1038/nmeth.4040>
- Rayleigh L (1907) **On our perception of sound direction** *Philos Mag* **13**:214–232
- Rupprecht P *et al.* (2021) **A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging** *Nat Neurosci* **24**:1324–1337 <https://doi.org/10.1038/s41593-021-00895-5>
- Sadeghi M, Zhai X, Stevenson IH, Escabí MA (2019) **A neural ensemble correlation code for sound category identification** *PLoS Biology* <https://doi.org/10.1371/journal.pbio.3000449>
- Schindelin J *et al.* (2012) **Fiji: An open-source platform for biological-image analysis** *Nat Methods* **9**:676–682 <https://doi.org/10.1038/nmeth.2019>
- Schnupp JW, King a J (1997) **Coding for auditory space in the nucleus of the brachium of the inferior colliculus in the ferret** *J Neurophysiol* **78**:2717–2731 <https://doi.org/10.1152/jn.1997.78.5.2717>
- Schofield BR, Beebe NL (2019) **Subtypes of GABAergic Cells in the Inferior Colliculus** *Hear Res* **376**:1–10 <https://doi.org/10.1016/j.heares.2018.10.001>
- Stringer C, Michaelos M, Tsyboulski D, Lindo SE, Pachitariu M (2021) **High-precision coding in visual cortex** *Cell* **184**:2767–2778 <https://doi.org/10.1016/j.cell.2021.03.042>
- Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) **Spontaneous behaviors drive multidimensional, brainwide activity** *Science* **364** <https://doi.org/10.1126/science.aav7893>

Vanwalleghe G, Constantin L, Scott EK (2021) **Calcium Imaging and the Curse of Negativity** *Front Neural Circuits* **14**:1–10 <https://doi.org/10.3389/fncir.2020.607391>

Weisenburger S, Prevedel R, Vaziri A (2017) **Quantitative evaluation of two-photon calcium imaging modalities for high-speed volumetric calcium imaging in scattering brain tissue** *bioRxiv* <https://doi.org/10.1101/115659>

Weisenburger S *et al.* (2019) **Volumetric Ca²⁺ Imaging in the Mouse Brain Using Hybrid Multiplexed Sculpted Light Microscopy** *Cell* **177**:1050–1066 <https://doi.org/10.1016/j.cell.2019.03.011>

Winer JA, Chernock ML, Larue DT, Cheung SW (2002) **Descending projections to the inferior colliculus from the posterior thalamus and the auditory cortex in rat, cat, and monkey** *Hear Res* **168**:181–195 [https://doi.org/10.1016/S0378-5955\(02\)00489-6](https://doi.org/10.1016/S0378-5955(02)00489-6)

Wong AB, Borst JGG (2019) **Tonotopic and non-auditory organization of the mouse dorsal inferior colliculus revealed by two-photon imaging** *Elife* **8**:1–50 <https://doi.org/10.7554/eLife.49091>

Zhou J, Shore S (2006) **Convergence of spinal trigeminal and cochlear nucleus projections in the inferior colliculus of the guinea pig** *J Comp Neurol* **495**:100–112 <https://doi.org/10.1002/cne.20863>

Editors

Reviewing Editor

Leopoldo Petreanu

Champalimaud Center for the Unknown, Lisbon, Portugal

Senior Editor

Barbara Shinn-Cunningham

Carnegie Mellon University, Pittsburgh, United States of America

Reviewer #1 (Public review):

Summary:

In this study, the authors address whether the dorsal nucleus of the inferior colliculus (DCIC) in mice encodes sound source location within the front horizontal plane (i.e., azimuth). They do this using volumetric two-photon Ca²⁺ imaging and high-density silicon probes (Neuropixels) to collect single-unit data. Such recordings are beneficial because they allow large populations of simultaneous neural data to be collected. Their main results and the claims about those results are the following:

- (1) DCIC single-unit responses have high trial-to-trial variability (i.e., neural noise);
- (2) approximately 32% to 40% of DCIC single units have responses that are sensitive to sound source azimuth;
- (3) single-trial population responses (i.e., the joint response across all sampled single units in an animal) encode sound source azimuth "effectively" (as stated in the title) in that localization decoding error matches average mouse discrimination thresholds;
- (4) DCIC can encode sound source azimuth in a similar format to that in the central nucleus of the inferior colliculus (as stated in the Abstract);
- (5) evidence of noise correlation between pairs of neurons exists;

and 6) noise correlations between responses of neurons help reduce population decoding error.

While simultaneous recordings are not necessary to demonstrate results #1, #2, and #4, they are necessary to demonstrate results #3, #5, and #6.

Strengths:

- Important research question to all researchers interested in sensory coding in the nervous system.
- State-of-the-art data collection: volumetric two-photon Ca²⁺ imaging and extracellular recording using high-density probes. Large neuronal data sets.
- Confirmation of imaging results (lower temporal resolution) with more traditional microelectrode results (higher temporal resolution).
- Clear and appropriate explanation of surgical and electrophysiological methods. I cannot comment on the appropriateness of the imaging methods.

Strength of evidence for the claims of the study:

(1) DCIC single-unit responses have high trial-to-trial variability -
The authors' data clearly shows this.

(2) Approximately 32% to 40% of DCIC single units have responses that are sensitive to sound source azimuth -

The sensitivity of each neuron's response to sound source azimuth was tested with a Kruskal-Wallis test, which is appropriate since response distributions were not normal. Using this statistical test, only 8% of neurons (median for imaging data) were found to be sensitive to azimuth, and the authors noted this was not significantly different than the false positive rate. The Kruskal-Wallis test was not reported for electrophysiological data. The authors suggested that low numbers of azimuth-sensitive units resulting from the statistical analysis may be due to the combination of high neural noise and relatively low number of trials, which would reduce statistical power of the test. This is likely true, and highlights a weakness in the experimental design (i.e., relatively small number of trials). The authors went on to perform a second test of azimuth sensitivity—a chi-squared test—and found 32% (imaging) and 40% (e-phys) of single units to have statistically significant sensitivity. However, the use of a chi-squared test is questionable because it is meant to be used between two categorical variables, and neural response had to be binned before applying the test.

(3) Single-trial population responses encode sound source azimuth "effectively" in that localization decoding error matches average mouse discrimination thresholds -

If only one neuron in a population had responses that were sensitive to azimuth, we would expect that decoding azimuth from observation of that one neuron's response would perform better than chance. By observing the responses of more than one neuron (if more than one were sensitive to azimuth), we would expect performance to increase. The authors found that decoding from the whole population response was no better than chance. They argue (reasonably) that this is because of overfitting of the decoder model—too few trials were used to fit too many parameters—and provide evidence from decoding combined with principal components analysis which suggests that overfitting is occurring. What is troubling is the performance of the decoder when using only a handful of "top-ranked" neurons (in terms of azimuth sensitivity) (Fig. 4F and G). Decoder performance seems to increase when going from one to two neurons, then decreases when going from two to three neurons, and doesn't get much better for more neurons than for one neuron alone. It seems likely there is more information about azimuth in the population response, but decoder performance is not able to capture it because spike count distributions in the decoder model are not being accurately estimated due to too few stimulus trials (14, on average). In other words, it seems likely that decoder performance is underestimating the ability of the DCIC population to encode sound source azimuth.

To get a sense of how effective a neural population is at coding a particular stimulus parameter, it is useful to compare population decoder performance to psychophysical performance. Unfortunately, mouse behavioral localization data do not exist. Instead, the authors compare decoder error to mouse left-right discrimination thresholds published previously by a different lab. However, this comparison is inappropriate because the decoder and the mice were performing different perceptual tasks. The decoder is classifying sound sources to 1 of 13 locations from left to right, whereas the mice were discriminating between left or right sources centered around zero degrees. The errors in these two tasks represent different things. The two data sets may potentially be more accurately compared by extracting information from the confusion matrices of population decoder performance. For example, when the stimulus was at -30 deg, how often did the decoder classify the stimulus to a lefthand azimuth? Likewise, when the stimulus was +30 deg, how often did the decoder classify the stimulus to a righthand azimuth?

(4) DCIC can encode sound source azimuth in a similar format to that in the central nucleus of the inferior colliculus -

It is unclear what exactly the authors mean by this statement in the Abstract. There are major differences in the encoding of azimuth between the two neighboring brain areas: a large majority of neurons in the CNIC are sensitive to azimuth (and strongly so), whereas the present study shows a minority of azimuth-sensitive neurons in the DCIC. Furthermore, CNIC neurons fire reliably to sound stimuli (low neural noise), whereas the present study shows that DCIC neurons fire more erratically (high neural noise).

(5) Evidence of noise correlation between pairs of neurons exists -

The authors' data and analyses seem appropriate and sufficient to justify this claim.

(6) Noise correlations between responses of neurons help reduce population decoding error -

The authors show convincing analysis that performance of their decoder increased when simultaneously measured responses were tested (which include noise correlation) than when scrambled-trial responses were tested (eliminating noise correlation). This makes it seem likely that noise correlation in the responses improved decoder performance. The authors mention that the naïve Bayesian classifier was used as their decoder for computational efficiency, presumably because it assumes no noise correlation and, therefore, assumes responses of individual neurons are independent of each other across trials to the same stimulus. The use of a decoder that assumes independence seems key here in testing the hypothesis that noise correlation contains information about sound source azimuth. The logic of using this decoder could be more clearly spelled out to the reader. For example, if the null hypothesis is that noise correlations do not carry azimuth information, then a decoder that assumes independence should perform the same whether population responses are simultaneous or scrambled. The authors' analysis showing a difference in performance between these two cases provides evidence against this null hypothesis.

Minor weakness:

- Most studies of neural encoding of sound source azimuth are done in a noise-free environment, but the experimental setup in the present study had substantial background noise. This complicates comparison of the azimuth tuning results in this study to those of other studies. One is left wondering if azimuth sensitivity would have been greater in the absence of background noise, particularly for the imaging data where the signal was only about 12 dB above the noise.

<https://doi.org/10.7554/eLife.97598.2.sa3>

Reviewer #2 (Public review):

In the present study, Boffi et al. investigate the manner in which the dorsal cortex of the of the inferior colliculus (DCIC), an auditory midbrain area, encodes sound location azimuth in awake, passively listening mice. By employing volumetric calcium imaging (scanned temporal focusing or s-TeFo), complemented with high-density electrode electrophysiological recordings (neuropixels probes), they show that sound-evoked responses are exquisitely noisy, with only a small portion of neurons (units) exhibiting spatial sensitivity. Nevertheless, a naïve Bayesian classifier was able to predict the presented azimuth based on the responses from small populations of these spatially sensitive units. A portion of the spatial information was provided by correlated trial-to-trial response variability between individual units (noise correlations). The study presents a novel characterization of spatial auditory coding in a non-canonical structure, representing a noteworthy contribution specifically to the auditory field and generally to systems neuroscience, due to its implementation of state-of-the-art techniques in an experimentally challenging brain region. However, nuances in the calcium imaging dataset and the naïve Bayesian classifier warrant caution when interpreting some of the results.

Strengths:

The primary strength of the study lies in its methodological achievements, which allowed the authors to collect a comprehensive and novel dataset. While the DCIC is a dorsal structure, it extends up to a millimetre in depth, making it optically challenging to access in its entirety. It is also more highly myelinated and vascularised compared to e.g., the cerebral cortex, compounding the problem. The authors successfully overcame these challenges and present an impressive volumetric calcium imaging dataset. Furthermore, they corroborated this dataset with electrophysiological recordings, which produced overlapping results. This methodological combination ameliorates the natural concerns that arise from inferring neuronal activity from calcium signals alone, which are in essence an indirect measurement thereof.

Another strength of the study is its interdisciplinary relevance. For the auditory field, it represents a significant contribution to the question of how auditory space is represented in the mammalian brain. "Space" per se is not mapped onto the basilar membrane of the cochlea and must be computed entirely within the brain. For azimuth, this requires the comparison between miniscule differences between the timing and intensity of sounds arriving at each ear. It is now generally thought that azimuth is initially encoded in two, opposing hemispheric channels, but the extent to which this initial arrangement is maintained throughout the auditory system remains an open question. The authors observe only a slight contralateral bias in their data, suggesting that sound source azimuth in the DCIC is encoded in a more nuanced manner compared to earlier processing stages of the auditory hindbrain. This is interesting because it is also known to be an auditory structure to receive more descending inputs from the cortex.

Systems neuroscience continues to strive for the perfection of imaging novel, less accessible brain regions. Volumetric calcium imaging is a promising emerging technique, allowing the simultaneous measurement of large populations of neurons in three dimensions. But this necessitates corroboration with other methods, such as electrophysiological recordings, which the authors achieve. The dataset moreover highlights the distinctive characteristics of neuronal auditory representations in the brain. Its signals can be exceptionally sparse and noisy, which provide an additional layer of complexity in the processing and analysis of such datasets. This will undoubtedly be useful for future studies of other less accessible structures with sparse responsiveness.

Weaknesses:

Although the primary finding that small populations of neurons carry enough spatial information for a naïve Bayesian classifier to reasonably decode the presented stimulus is not called into question, certain idiosyncrasies, in particular the calcium imaging dataset and model, complicate specific interpretations of the model output, and the readership is urged to interpret these aspects of the study's conclusions with caution.

I remain in favour of volumetric calcium imaging as a suitable technique for the study, but the presently constrained spatial resolution is insufficient to unequivocally identify regions of interest as cell bodies (and are instead referred to as "units" akin to those of electrophysiological recordings). It remains possible that the imaging set is inadvertently influenced by non-somatic structures (including neuropil), which could report neuronal activity differently than cell bodies. Due to the lack of a comprehensive ground-truth comparison in this regard (which to my knowledge is impossible to achieve with current technology), it is difficult to imagine how many informative such units might have been missed because their signals were influenced by spurious, non-somatic signals, which could have subsequently misled the models. The authors reference the original Nature Methods article (Prevedel et al., 2016) throughout the manuscript, presumably in order to avoid having to repeat previously published experimental metrics. But the DCIC is neither the cortex nor hippocampus (for which the method was originally developed) and may not have the same light scattering properties (not to mention neuronal noise levels). Although the corroborative electrophysiology data largely alleviates these concerns for this particular study, the readership should be cognisant of such caveats, in particular those who are interested in implementing the technique for their own research.

A related technical limitation of the calcium imaging dataset is the relatively low number of trials (14) given the inherently high level of noise (both neuronal and imaging). Volumetric calcium imaging, while offering a uniquely expansive field of view, requires relatively high average excitation laser power (in this case nearly 200 mW), a level of exposure the authors may have wanted to minimise by maintaining a low number of repetitions, but I yield to them to explain. Calcium imaging is also inherently slow, requiring relatively long inter-stimulus intervals (in this case 5 s). This unfortunately renders any model designed to predict a stimulus (in this case sound azimuth) from particularly noisy population neuronal data like these as highly prone to overfitting, to which the authors correctly admit after a model trained on the entire raw dataset failed to perform significantly above chance level. This prompted them to feed the model only with data from neurons with the highest spatial sensitivity. This ultimately produced reasonable performance (and was implemented throughout the rest of the study), but it remains possible that if the model was fed with more repetitions of imaging data, its performance would have been more stable across the number of units used to train it. (All models trained with imaging data eventually failed to converge.) However, I also see these limitations as an opportunity to improve the technology further, which I reiterate will be generally important for volume imaging of other sparse or noisy calcium signals in the brain.

Indeed, in separate comments to these remarks, the authors confirmed that the low number of trials was technically limited, to which I emphasise is to no fault of their own. However, they also do not report this as a typical imaging constraint, such as photobleaching, but rather because the animals exhibited signs of stress and discomfort at longer imaging periods. From an animal welfare perspective, I would encourage the authors to state this in the methods for transparency. It would demonstrate their adherence to animal welfare policies, which I find to be an incredibly strong argument for limiting the number of trials in their study.

Transitioning to the naïve Bayesian classifier itself, I first openly ask the authors to justify their choice of this specific model. There are countless types of classifiers for these data, each with their own pros and cons. Did they actually try other models (such as support vector

machines), which ultimately failed? If so, these negative results (even if mentioned en passant) would be extremely valuable to the community, in my view. I ask this specifically because different methods assume correspondingly different statistical properties of the input data, and to my knowledge naïve Bayesian classifiers assume that predictors (neuronal responses) are assumed to be independent within a class (azimuth). As the authors show that noise correlations are informative in predicting azimuth, I wonder why they chose a model that doesn't take advantage of these statistical regularities. It could be because of technical considerations (they mention computing efficiency), but I am left generally uncertain about the specific logic that was used to guide the authors through their analytical journey.

In a revised version of the manuscript, the authors indeed justify their choice of the naïve Bayesian classifier as a conservative approach (not taking into account noise correlations), which could only improve with other models (that do). They even tested various other commonly used models, such as support vector machines and k-nearest neighbours, to name a few, but do not report these efforts in the main manuscript. Interestingly, these models, which I supposed would perform better in fact did not overall - a finding that I have no way of interpreting but nevertheless find interesting. I would thus encourage the authors to include these results in a figure supplement and mention it en passant while justifying their selection of model (but please include detailed model parameters in the methods section).

That aside, there remain other peculiarities in model performance that warrant further investigation. For example, what spurious features (or lack of informative features) in these additional units prevented the models of imaging data from converging? In an orthogonal question, did the most spatially sensitive units share any detectable tuning features? A different model trained with electrophysiology data in contrast did not collapse in the range of top-ranked units plotted. Did this model collapse at some point after adding enough units, and how well did that correlate with the model for the imaging data? How well did the form (and diversity) of the spatial tuning functions as recorded with electrophysiology resemble their calcium imaging counterparts? These fundamental questions could be addressed with more basic, but transparent analyses of the data (e.g., the diversity of spatial tuning functions of their recorded units across the population). Even if the model extracts features that are not obvious to the human eye in traditional visualisations, I would still find this interesting.

Although these questions were not specifically addressed in the revised version of the manuscript, I also admit that I did not intend to assert that these should necessarily fall within the scope of the present study. I rather posed them as hypothetical directions one could pursue in future studies. Finally, further concerns I had with statements regarding the physiological meaning of the findings have been ameliorated by nicely modified statements, thus bringing transparency to the readership, which I appreciate.

In summary, the present study represents a significant body of work that contributes substantially to the field of spatial auditory coding and systems neuroscience. However, limitations of the imaging dataset and model as applied in the study muddles concrete conclusions about how the DCIC precisely encodes sound source azimuth and even more so to sound localisation in a behaving animal. Nevertheless, it presents a novel and unique dataset, which, regardless of secondary interpretation, corroborates the general notion that auditory space is encoded in an extraordinarily complex manner in the mammalian brain.

<https://doi.org/10.7554/eLife.97598.2.sa2>

Reviewer #3 (Public review):

Summary:

Boffi and colleagues sought to quantify the single-trial, azimuthal information in the dorsal cortex of the inferior colliculus (DCIC), a relatively understudied subnucleus of the auditory midbrain. They accomplished this by using two complementary recording methods while mice passively listened to sounds at different locations: calcium imaging that recorded large neuronal populations but with poor temporal precision and multi-contact electrode arrays that recorded smaller neuronal populations with exact temporal precision. DCIC neurons respond variably, with inconsistent activity to sound onset and complex azimuthal tuning. Some of this variability was explained by ongoing head movements. The authors used a naïve Bayes decoder to probe the azimuthal information contained in the response of DCIC neurons on single trials. The decoder failed to classify sound location better than chance when using the raw population responses but performed significantly better than chance when using the top principal components of the population. Units with the most azimuthal tuning were distributed throughout the DCIC, possessed contralateral bias, and positively correlated responses. Interestingly, inter-trial shuffling decreased decoding performance, indicating that noise correlations contributed to decoder performance. Overall, Boffi and colleagues, quantified the azimuthal information available in the DCIC while mice passively listened to sounds, a first step in evaluating if and how the DCIC could contribute to sound localization.

Strengths:

The authors should be commended for collection of this dataset. When done in isolation (which is typical), calcium imaging and linear array recordings have intrinsic weaknesses. However, those weaknesses are alleviated when done in conjunction - especially when the data is consistent. This data set is extremely rich and will be of use for those interested in auditory midbrain responses to variable sound locations, correlations with head movements, and neural coding.

The DCIC neural responses are complex with variable responses to sound onset, complex azimuthal tuning and large inter-sound interval responses. Nonetheless, the authors do a decent job in wrangling these complex responses: finding non-canonical ways of determining dependence on azimuth and using interpretable decoders to extract information from the population.

Weaknesses:

The decoding results are a bit strange, likely because the population response is quite noisy on any given trial. Raw population responses failed to provide sufficient information concerning azimuth for significant decoding. Importantly, the decoder performed better than chance when certain principal components or top ranked units contributed but did not saturate with the addition of components or top ranked units. So, although there is azimuthal information in the recorded DCIC populations - azimuthal information appears somewhat difficult to extract.

Although necessary given the challenges associated with sampling many conditions with technically difficult recording methods, the limited number of stimulus repeats precludes interpretable characterization of the heterogeneity across the population. Nevertheless, the dataset is public so those interested can explore the diversity of the responses.

The observations from Boffi and colleagues raises the question: what drives neurons in the DCIC to respond? Sound azimuth appears to be a small aspect of the DCIC response. For example, the first 20 principal components which explain roughly 80% of the response variance are insufficient input for the decoder to predict sound azimuth above chance. Furthermore, snout and ear movements correlate with the population response in the DCIC (the ear movements are particularly peculiar given they seem to predict sound presentation). Other movements may be of particular interest to control for (e.g. eye movements are known to interact with IC responses in the primate). These observations, along with reported

variance to sound onsets and inter-sound intervals, question the impact of azimuthal information emerging from DCIC responses. This is certainly out of scope for any one singular study to answer, but, hopefully, future work will elucidate the dominant signals in the DCIC population. It may be intuitive that engagement in a sound localization task may push azimuthal signals to the forefront of DCIC response, but azimuthal information could also easily be overtaken by other signals (e.g. movement, learning).

Boffi and colleagues set out to parse the azimuthal information available in the DCIC on a single trial. They largely accomplish this goal and are able to extract this information when allowing the units that contain more information about sound location to contribute to their decoding (e.g., through PCA or decoding on their activity specifically). Interestingly, they also found that positive noise correlations between units with similar azimuthal preferences facilitate this decoding - which is unusual given that this is typically thought to limit information. The dataset will be of value to those interested in the DCIC and to anyone interested in the role of noise correlations in population coding. Although this work is first step into parsing the information available in the DCIC, it remains difficult to interpret if/how this azimuthal information is used in localization behaviors of engaged mice.

<https://doi.org/10.7554/eLife.97598.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

In this study, the authors address whether the dorsal nucleus of the inferior colliculus (DCIC) in mice encodes sound source location within the front horizontal plane (i.e., azimuth). They do this using volumetric two-photon Ca²⁺ imaging and high-density silicon probes (Neuropixels) to collect single-unit data. Such recordings are beneficial because they allow large populations of simultaneous neural data to be collected. Their main results and the claims about those results are the following:

- (1) DCIC single-unit responses have high trial-to-trial variability (i.e., neural noise);*
- (2) approximately 32% to 40% of DCIC single units have responses that are sensitive to sound source azimuth;*
- (3) single-trial population responses (i.e., the joint response across all sampled single units in an animal) encode sound source azimuth "effectively" (as stated in title) in that localization decoding error matches average mouse discrimination thresholds;*
- (4) DCIC can encode sound source azimuth in a similar format to that in the central nucleus of the inferior colliculus (as stated in Abstract);*
- (5) evidence of noise correlation between pairs of neurons exists;*
- and 6) noise correlations between responses of neurons help reduce population decoding error.*

While simultaneous recordings are not necessary to demonstrate results #1, #2, and #4, they are necessary to demonstrate results #3, #5, and #6.

Strengths:

- Important research question to all researchers interested in sensory coding in the nervous system.

- State-of-the-art data collection: volumetric two-photon Ca²⁺ imaging and extracellular recording using high-density probes. Large neuronal data sets.

- Confirmation of imaging results (lower temporal resolution) with more traditional microelectrode results (higher temporal resolution).

- Clear and appropriate explanation of surgical and electrophysiological methods. I cannot comment on the appropriateness of the imaging methods.

Strength of evidence for claims of the study:

(1) DCIC single-unit responses have high trial-to-trial variability - The authors' data clearly shows this.

(2) Approximately 32% to 40% of DCIC single units have responses that are sensitive to sound source azimuth - The sensitivity of each neuron's response to sound source azimuth was tested with a Kruskal-Wallis test, which is appropriate since response distributions were not normal. Using this statistical test, only 8% of neurons (median for imaging data) were found to be sensitive to azimuth, and the authors noted this was not significantly different than the false positive rate. The Kruskal-Wallis test was not performed on electrophysiological data. The authors suggested that low numbers of azimuth-sensitive units resulting from the statistical analysis may be due to the combination of high neural noise and relatively low number of trials, which would reduce statistical power of the test. This may be true, but if single-unit responses were moderately or strongly sensitive to azimuth, one would expect them to pass the test even with relatively low statistical power. At best, if their statistical test missed some azimuth-sensitive units, they were likely only weakly sensitive to azimuth. The authors went on to perform a second test of azimuth sensitivity—a chi-squared test—and found 32% (imaging) and 40% (e-phys) of single units to have statistically significant sensitivity. This feels a bit like fishing for a lower p-value. The Kruskal-Wallis test should have been left as the only analysis. Moreover, the use of a chi-squared test is questionable because it is meant to be used between two categorical variables, and neural response had to be binned before applying the test.

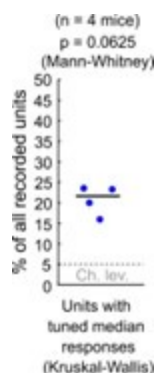
The determination of what is a physiologically relevant “moderate or strong azimuth sensitivity” is not trivial, particularly when comparing tuning across different relays of the auditory pathway like the CNIC, auditory cortex, or in our case DCIC, where physiologically relevant azimuth sensitivities might be different. This is likely the reason why azimuth sensitivity has been defined in diverse ways across the bibliography (see Groh, Kelly & Underhill, 2003 for an early discussion of this issue). These diverse approaches include reaching a certain percentage of maximal response modulation, like used by Day et al. (2012, 2015, 2016) in CNIC, and ANOVA tests, like used by Panniello et al. (2018) and Groh, Kelly & Underhill (2003) in auditory cortex and IC respectively. Moreover, the influence of response variability and biases in response distribution estimation due to limited sampling has not been usually accounted for in the determination of azimuth sensitivity.

As Reviewer #1 points out, in our study we used an appropriate ANOVA test (KruskalWallis) as a starting point to study response sensitivity to stimulus azimuth at DCIC. Please note that the $\alpha = 0.05$ used for this test is not based on experimental evidence about physiologically relevant azimuth sensitivity but instead is an arbitrary p-value threshold. Using this test on

the electrophysiological data, we found that ~ 21% of the simultaneously recorded single units reached significance ($n = 4$ mice). Nevertheless these percentages, in our small sample size ($n = 4$) were not significantly different from our false positive detection rate ($p = 0.0625$, Mann-Whitney, See Author response image 1 below). In consequence, for both our imaging (Fig. 3C) and electrophysiological data, we could not ascertain if the percentage of neurons reaching significance in these ANOVA tests were indeed meaningfully sensitive to azimuth or this was due to chance.

Author response image 1.

Percentage of the neuropixels recorded DCIC single units across mice that showed significant median response tuning, compared to false positive detection rate ($\alpha = 0.05$, chance level).



We reasoned that the observed markedly variable responses from DCIC units, which frequently failed to respond in many trials (Fig. 3D, 4A), in combination with the limited number of trial repetitions we could collect, results in under-sampled response distribution estimations. This under-sampling can bias the determination of stochastic dominance across azimuth response samples in Kruskal-Wallis tests. We would like to highlight that we decided not to implement resampling strategies to artificially increase the azimuth response sample sizes with “virtual trials”, in order to avoid “fishing for a smaller p-value”, when our collected samples might not accurately reflect the actual response population variability.

As an alternative to hypothesis testing based on ranking and determining stochastic dominance of one or more azimuth response samples (Kruskal-Wallis test), we evaluated the overall statistical dependency to stimulus azimuth of the collected responses. To do this we implement the Chi-square test by binning neuronal responses into categories. Binning responses into categories can reduce the influence of response variability to some extent, which constitutes an advantage of the Chi-square approach, but we note the important consideration that these response categories are arbitrary.

Altogether, we acknowledge that our Chi-square approach to define azimuth sensitivity is not free of limitations and despite enabling the interrogation of azimuth sensitivity at DCIC, its interpretability might not extend to other brain regions like CNIC or auditory cortex. Nevertheless we hope the aforementioned arguments justify why the Kruskal-Wallis test simply could not “have been left as the only analysis”.

(3) Single-trial population responses encode sound source azimuth "effectively" in that localization decoding error matches average mouse discrimination thresholds - If only one neuron in a population had responses that were sensitive to azimuth, we would expect that decoding azimuth from observation of that one neuron's response would perform better than chance. By observing the responses of more than one neuron (if more than one were sensitive to azimuth), we would expect performance to increase. The

authors found that decoding from the whole population response was no better than chance. They argue (reasonably) that this is because of overfitting of the decoder model too few trials used to fit too many parameters-and provide evidence from decoding combined with principal components analysis which suggests that overfitting is occurring. What is troubling is the performance of the decoder when using only a handful of "topranked" neurons (in terms of azimuth sensitivity) (Fig. 4F and G). Decoder performance seems to increase when going from one to two neurons, then decreases when going from two to three neurons, and doesn't get much better for more neurons than for one neuron alone. It seems likely there is more information about azimuth in the population response, but decoder performance is not able to capture it because spike count distributions in the decoder model are not being accurately estimated due to too few stimulus trials (14, on average). In other words, it seems likely that decoder performance is underestimating the ability of the DCIC population to encode sound source azimuth.

To get a sense of how effective a neural population is at coding a particular stimulus parameter, it is useful to compare population decoder performance to psychophysical performance. Unfortunately, mouse behavioral localization data do not exist. Therefore, the authors compare decoder error to mouse left-right discrimination thresholds published previously by a different lab. However, this comparison is inappropriate because the decoder and the mice were performing different perceptual tasks. The decoder is classifying sound sources to 1 of 13 locations from left to right, whereas the mice were discriminating between left or right sources centered around zero degrees. The errors in these two tasks represent different things. The two data sets may potentially be more accurately compared by extracting information from the confusion matrices of population decoder performance. For example, when the stimulus was at -30 deg, how often did the decoder classify the stimulus to a lefthand azimuth? Likewise, when the stimulus was +30 deg, how often did the decoder classify the stimulus to a righthand azimuth?

The azimuth discrimination error reported by Lauer et al. (2011) comes from engaged and highly trained mice, which is a very different context to our experimental setting with untrained mice passively listening to stimuli from 13 random azimuths. Therefore we did not perform analyses or interpretations of our results based on the behavioral task from Lauer et al. (2011) and only made the qualitative observation that the errors match for discussion.

We believe it is further important to clarify that Lauer et al. (2011) tested the ability of mice to discriminate between a positively conditioned stimulus (reference speaker at 0° center azimuth associated to a liquid reward) and a negatively conditioned stimulus (coming from one of five comparison speakers positioned at 20°, 30°, 50°, 70 and 90° azimuth, associated to an electrified lickport) in a conditioned avoidance task. In this task, mice are not precisely "discriminating between left or right sources centered around zero degrees", making further analyses to compare the experimental design of Lauer et al (2011) and ours even more challenging for valid interpretation.

(4) DCIC can encode sound source azimuth in a similar format to that in the central nucleus of the inferior colliculus - It is unclear what exactly the authors mean by this statement in the Abstract. There are major differences in the encoding of azimuth between the two neighboring brain areas: a large majority of neurons in the CNIC are sensitive to azimuth (and strongly so), whereas the present study shows a minority of azimuth-sensitive neurons in the DCIC. Furthermore, CNIC neurons fire reliably to sound stimuli (low neural noise), whereas the present study shows that DCIC neurons fire more erratically (high neural noise).

Since sound source azimuth is reported to be encoded by population activity patterns at CNIC (Day and Delgutte, 2013), we refer to a population activity pattern code as the “similar format” in which this information is encoded at DCIC. Please note that this is a qualitative comparison and we do not claim this is the “same format”, due to the differences the reviewer precisely describes in the encoding of azimuth at CNIC where a much larger majority of neurons show stronger azimuth sensitivity and response reliability with respect to our observations at DCIC. By this qualitative similarity of encoding format we specifically mean the similar occurrence of activity patterns from azimuth sensitive subpopulations of neurons in both CNIC and DCIC, which carry sufficient information about the stimulus azimuth for a sufficiently accurate prediction with regard to the behavioral discrimination ability.

(5) Evidence of noise correlation between pairs of neurons exists - The authors' data and analyses seem appropriate and sufficient to justify this claim.

(6) Noise correlations between responses of neurons help reduce population decoding error - The authors show convincing analysis that performance of their decoder increased when simultaneously measured responses were tested (which include noise correlation) than when scrambled-trial responses were tested (eliminating noise correlation). This makes it seem likely that noise correlation in the responses improved decoder performance. The authors mention that the naïve Bayesian classifier was used as their decoder for computational efficiency, presumably because it assumes no noise correlation and, therefore, assumes responses of individual neurons are independent of each other across trials to the same stimulus. The use of decoder that assumes independence seems key here in testing the hypothesis that noise correlation contains information about sound source azimuth. The logic of using this decoder could be more clearly spelled out to the reader. For example, if the null hypothesis is that noise correlations do not carry azimuth information, then a decoder that assumes independence should perform the same whether population responses are simultaneous or scrambled. The authors' analysis showing a difference in performance between these two cases provides evidence against this null hypothesis.

We sincerely thank the reviewer for this careful and detailed consideration of our analysis approach. Following the reviewer's constructive suggestion, we justified the decoder choice in the results section at the last paragraph of page 18:

“To characterize how the observed positive noise correlations could affect the representation of stimulus azimuth by DCIC top ranked unit population responses, we compared the decoding performance obtained by classifying the single-trial response patterns from top ranked units in the modeled decorrelated datasets versus the acquired data (with noise correlations). With the intention to characterize this with a conservative approach that would be less likely to find a contribution of noise correlations as it assumes response independence, we relied on the naïve Bayes classifier for decoding throughout the study. Using this classifier, we observed that the modeled decorrelated datasets produced stimulus azimuth prediction error distributions that were significantly shifted towards higher decoding errors (Fig. 5B, C) and, in our imaging datasets, were not significantly different from chance level (Fig. 5B). Altogether, these results suggest that the detected noise correlations in our simultaneously acquired datasets can help reduce the error of the IC population code for sound azimuth.”

Minor weakness:

- Most studies of neural encoding of sound source azimuth are done in a noise-free environment, but the experimental setup in the present study had substantial

background noise. This complicates comparison of the azimuth tuning results in this study to those of other studies. One is left wondering if azimuth sensitivity would have been greater in the absence of background noise, particularly for the imaging data where the signal was only about 12 dB above the noise. The description of the noise level and signal + noise level in the Methods should be made clearer. Mice hear from about 2.5 - 80 kHz, so it is important to know the noise level within this band as well as specifically within the band overlapping with the signal.

We agree with the reviewer that this information is useful. In our study, the background R.M.S. SPL during imaging across the mouse hearing range (2.5-80kHz) was 44.53 dB and for neuropixels recordings 34.68 dB. We have added this information to the methods section of the revised manuscript.

Reviewer #2 (Public Review):

In the present study, Boffi et al. investigate the manner in which the dorsal cortex of the inferior colliculus (DCIC), an auditory midbrain area, encodes sound location azimuth in awake, passively listening mice. By employing volumetric calcium imaging (scanned temporal focusing or s-TeFo), complemented with high-density electrode electrophysiological recordings (neuropixels probes), they show that sound-evoked responses are exquisitely noisy, with only a small portion of neurons (units) exhibiting spatial sensitivity. Nevertheless, a naïve Bayesian classifier was able to predict the presented azimuth based on the responses from small populations of these spatially sensitive units. A portion of the spatial information was provided by correlated trial-to-trial response variability between individual units (noise correlations). The study presents a novel characterization of spatial auditory coding in a non-canonical structure, representing a noteworthy contribution specifically to the auditory field and generally to systems neuroscience, due to its implementation of state-of-the-art techniques in an experimentally challenging brain region. However, nuances in the calcium imaging dataset and the naïve Bayesian classifier warrant caution when interpreting some of the results.

Strengths:

The primary strength of the study lies in its methodological achievements, which allowed the authors to collect a comprehensive and novel dataset. While the DCIC is a dorsal structure, it extends up to a millimetre in depth, making it optically challenging to access in its entirety. It is also more highly myelinated and vascularised compared to e.g., the cerebral cortex, compounding the problem. The authors successfully overcame these challenges and present an impressive volumetric calcium imaging dataset. Furthermore, they corroborated this dataset with electrophysiological recordings, which produced overlapping results. This methodological combination ameliorates the natural concerns that arise from inferring neuronal activity from calcium signals alone, which are in essence an indirect measurement thereof.

Another strength of the study is its interdisciplinary relevance. For the auditory field, it represents a significant contribution to the question of how auditory space is represented in the mammalian brain. "Space" per se is not mapped onto the basilar membrane of the cochlea and must be computed entirely within the brain. For azimuth, this requires the comparison between miniscule differences between the timing and intensity of sounds arriving at each ear. It is now generally thought that azimuth is initially encoded in two, opposing hemispheric channels, but the extent to which this initial arrangement is maintained throughout the auditory system remains an open question. The authors observe only a slight contralateral bias in their data, suggesting that sound source azimuth in the DCIC is encoded in a more nuanced manner compared

to earlier processing stages of the auditory hindbrain. This is interesting, because it is also known to be an auditory structure to receive more descending inputs from the cortex.

Systems neuroscience continues to strive for the perfection of imaging novel, less accessible brain regions. Volumetric calcium imaging is a promising emerging technique, allowing the simultaneous measurement of large populations of neurons in three dimensions. But this necessitates corroboration with other methods, such as electrophysiological recordings, which the authors achieve. The dataset moreover highlights the distinctive characteristics of neuronal auditory representations in the brain. Its signals can be exceptionally sparse and noisy, which provide an additional layer of complexity in the processing and analysis of such datasets. This will be undoubtedly useful for future studies of other less accessible structures with sparse responsiveness.

Weaknesses:

Although the primary finding that small populations of neurons carry enough spatial information for a naïve Bayesian classifier to reasonably decode the presented stimulus is not called into question, certain idiosyncrasies, in particular the calcium imaging dataset and model, complicate specific interpretations of the model output, and the readership is urged to interpret these aspects of the study's conclusions with caution.

I remain in favour of volumetric calcium imaging as a suitable technique for the study, but the presently constrained spatial resolution is insufficient to unequivocally identify regions of interest as cell bodies (and are instead referred to as "units" akin to those of electrophysiological recordings). It remains possible that the imaging set is inadvertently influenced by non-somatic structures (including neuropil), which could report neuronal activity differently than cell bodies. Due to the lack of a comprehensive ground-truth comparison in this regard (which to my knowledge is impossible to achieve with current technology), it is difficult to imagine how many informative such units might have been missed because their signals were influenced by spurious, non-somatic signals, which could have subsequently misled the models. The authors reference the original Nature Methods article (Prevedel et al., 2016) throughout the manuscript, presumably in order to avoid having to repeat previously published experimental metrics. But the DCIC is neither the cortex nor hippocampus (for which the method was originally developed) and may not have the same light scattering properties (not to mention neuronal noise levels). Although the corroborative electrophysiology data largely alleviates these concerns for this particular study, the readership should be cognisant of such caveats, in particular those who are interested in implementing the technique for their own research.

A related technical limitation of the calcium imaging dataset is the relatively low number of trials (14) given the inherently high level of noise (both neuronal and imaging). Volumetric calcium imaging, while offering a uniquely expansive field of view, requires relatively high average excitation laser power (in this case nearly 200 mW), a level of exposure the authors may have wanted to minimise by maintaining a low the number of repetitions, but I yield to them to explain.

We assumed that the levels of heating by excitation light measured at the neocortex in Prevedel et al. (2016), were representative for DCIC also. Nevertheless, we recognize this approximation might not be very accurate, due to the differences in tissue architecture and vascularization from these two brain areas, just to name a few factors. The limiting factor preventing us from collecting more trials in our imaging sessions was that we observed signs of discomfort or slight distress in some mice after ~30 min of imaging in our custom setup, which we established as a humane end point to prevent distress. In consequence imaging sessions were kept to 25 min in duration, limiting the number of trials collected. However we

cannot rule out that with more extensive habituation prior to experiments the imaging sessions could be prolonged without these signs of discomfort or if indeed influence from our custom setup like potential heating of the brain by illumination light might be the causing factor of the observed distress. Nevertheless, we note that previous work has shown that ~200mW average power is a safe regime for imaging in the cortex by keeping brain heating minimal (Prevedel et al., 2016), without producing the lasting damages observed by immunohistochemistry against apoptosis markers above 250mW (Podgorski and Ranganathan 2016, <https://doi.org/10.1152/jn.00275.2016>).

Calcium imaging is also inherently slow, requiring relatively long inter-stimulus intervals (in this case 5 s). This unfortunately renders any model designed to predict a stimulus (in this case sound azimuth) from particularly noisy population neuronal data like these as highly prone to overfitting, to which the authors correctly admit after a model trained on the entire raw dataset failed to perform significantly above chance level. This prompted them to feed the model only with data from neurons with the highest spatial sensitivity. This ultimately produced reasonable performance (and was implemented throughout the rest of the study), but it remains possible that if the model was fed with more repetitions of imaging data, its performance would have been more stable across the number of units used to train it. (All models trained with imaging data eventually failed to converge.) However, I also see these limitations as an opportunity to improve the technology further, which I reiterate will be generally important for volume imaging of other sparse or noisy calcium signals in the brain.

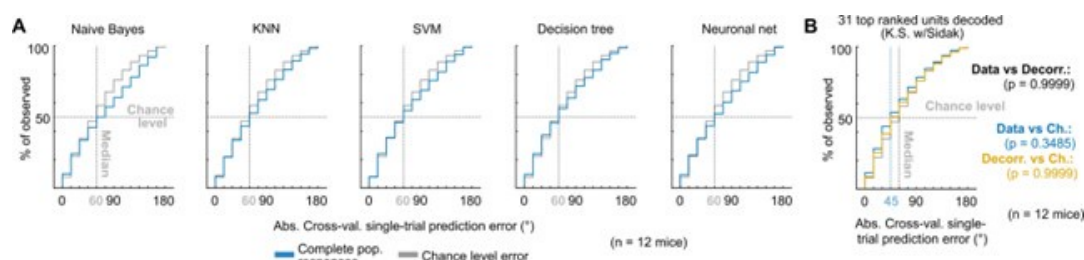
Transitioning to the naïve Bayesian classifier itself, I first openly ask the authors to justify their choice of this specific model. There are countless types of classifiers for these data, each with their own pros and cons. Did they actually try other models (such as support vector machines), which ultimately failed? If so, these negative results (even if mentioned en passant) would be extremely valuable to the community, in my view. I ask this specifically because different methods assume correspondingly different statistical properties of the input data, and to my knowledge naïve Bayesian classifiers assume that predictors (neuronal responses) are assumed to be independent within a class (azimuth). As the authors show that noise correlations are informative in predicting azimuth, I wonder why they chose a model that doesn't take advantage of these statistical regularities. It could be because of technical considerations (they mention computing efficiency), but I am left generally uncertain about the specific logic that was used to guide the authors through their analytical journey.

One of the main reasons we chose the naïve Bayesian classifier is indeed because it assumes that the responses of the simultaneously recorded neurons are independent and therefore it does not assume a contribution of noise correlations to the estimation of the posterior probability of each azimuth. This model would represent the null hypothesis that noise correlations do not contribute to the encoding of stimulus azimuth, which would be verified by an equal decoding outcome from correlated or decorrelated datasets. Since we observed that this is not the case, the model supports the alternative hypothesis that noise correlations do indeed influence stimulus azimuth encoding. We wanted to test these hypotheses with the most conservative approach possible that would be least likely to find a contribution of noise correlations. Other relevant reasons that justify our choice of the naïve Bayesian classifier are its robustness against the limited numbers of trials we could collect in comparison to other more “data hungry” classifiers like SVM, KNN, or artificial neuronal nets. We did perform preliminary tests with alternative classifiers but the obtained decoding errors were similar when decoding the whole population activity (Author response image 2A). Dimensionality reduction following the approach described in the manuscript showed a tendency towards smaller decoding errors observed with an alternative classifier like KNN, but these errors were still larger than the ones observed with the naïve Bayesian classifier

(median error 45°). Nevertheless, we also observe a similar tendency for slightly larger decoding errors in the absence of noise correlations (decorrelated, Author response image 2B). Sentences detailing the logic of classifier choice are now included in the results section at page 10 and at the last paragraph of page 18 (see responses to Reviewer 1).

Author response image 2.

A) Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained using different classifiers (blue; KNN: K-nearest neighbors; SVM: support vector machine ensemble) and chance level distribution (gray) on the complete populations of imaged units. Cumulative distribution plots of the absolute cross-validated single-trial prediction errors obtained using a Bayes classifier (naive approximation for computation efficiency) to decode the single-trial response patterns from the 31 top ranked units in the simultaneously imaged datasets across mice (cyan), modeled decorrelated datasets (orange) and the chance level distribution associated with our stimulation paradigm (gray). Vertical dashed lines show the medians of cumulative distributions. K.S. w/Sidak: Kolmogorov-Smirnov with Sidak.



That aside, there remain other peculiarities in model performance that warrant further investigation. For example, what spurious features (or lack of informative features) in these additional units prevented the models of imaging data from converging?

Considering the amount of variability observed throughout the neuronal responses both in imaging and neuropixels datasets, it is easy to suspect that the information about stimulus azimuth carried in different amounts by individual DCIC neurons can be mixed up with information about other factors (Stringer et al., 2019). In an attempt to study the origin of these features that could confound stimulus azimuth decoding we explored their relation to face movement (Supplemental Figure 2), finding a correlation to snout movements, in line with previous work by Stringer et al. (2019).

In an orthogonal question, did the most spatially sensitive units share any detectable tuning features? A different model trained with electrophysiology data in contrast did not collapse in the range of top-ranked units plotted. Did this model collapse at some point after adding enough units, and how well did that correlate with the model for the imaging data?

Our electrophysiology datasets were much smaller in size (number of simultaneously recorded neurons) compared to our volumetric calcium imaging datasets, resulting in a much smaller total number of top ranked units detected per dataset. This precluded the determination of a collapse of decoder performance due to overfitting beyond the range plotted in Fig 4G.

How well did the form (and diversity) of the spatial tuning functions as recorded with electrophysiology resemble their calcium imaging counterparts? These fundamental questions could be addressed with more basic, but transparent analyses of the data (e.g.,

the diversity of spatial tuning functions of their recorded units across the population). Even if the model extracts features that are not obvious to the human eye in traditional visualisations, I would still find this interesting.

The diversity of the azimuth tuning curves recorded with calcium imaging (Fig. 3B) was qualitatively larger than the ones recorded with electrophysiology (Fig. 4B), potentially due to the larger sampling obtained with volumetric imaging. We did not perform a detailed comparison of the form and a more quantitative comparison of the diversity of these functions because the signals compared are quite different, as calcium indicator signal is subject to non linearities due to Ca^{2+} binding cooperativity and low pass filtering due to binding kinetics. We feared this could lead to misleading interpretations about the similarities or differences between the azimuth tuning functions in imaged and electrophysiology datasets. Our model uses statistical response dependency to stimulus azimuth, which does not rely on features from a descriptive statistic like mean response tuning. In this context, visualizing the trial-to-trial responses as a function of azimuth shows “features that are not obvious to the human eye in traditional visualizations” (Fig. 3D, left inset).

Finally, the readership is encouraged to interpret certain statements by the authors in the current version conservatively. How the brain ultimately extracts spatial neuronal data for perception is anyone's guess, but it is important to remember that this study only shows that a naïve Bayesian classifier could decode this information, and it remains entirely unclear whether the brain does this as well. For example, the model is able to achieve a prediction error that corresponds to the psychophysical threshold in mice performing a discrimination task ($\sim 30^\circ$ {degree sign}). Although this is an interesting coincidental observation, it does not mean that the two metrics are necessarily related. The authors correctly do not explicitly claim this, but the manner in which the prose flows may lead a non-expert into drawing that conclusion.

To avoid misleading the non-expert readers, we have clarified in the manuscript that the observed correspondence between decoding error and psychophysical threshold is explicitly coincidental.

Page 13, end of middle paragraph:

“If we consider the median of the prediction error distribution as an overall measure of decoding performance, the single-trial response patterns from subsamples of at least the 7 top ranked units produced median decoding errors that coincidentally matched the reported azimuth discrimination ability of mice (Fig 4G, minimum audible angle = 31°) (Lauer et al., 2011).”

Page 14, bottom paragraph:

“Decoding analysis (Fig. 4F) of the population response patterns from azimuth dependent top ranked units simultaneously recorded with neuropixels probes showed that the 4 top ranked units are the smallest subsample necessary to produce a significant decoding performance that coincidentally matches the discrimination ability of mice (31° (Lauer et al., 2011)) (Fig. 5F, G).”

We also added to the Discussion sentences clarifying that a relationship between these two variables remains to be determined and it also remains to be determined if the DCIC indeed performs a bayesian decoding computation for sound localization.

Page 20, bottom:

“... Concretely, we show that sound location coding does indeed occur at DCIC on the single trial basis, and that this follows a comparable mechanism to the characterized population code at CNIC (Day and Delgutte, 2013). However, it remains to be determined if indeed the DCIC network is physiologically capable of Bayesian decoding computations. Interestingly, the small number of DCIC top ranked units necessary to effectively decode stimulus azimuth suggests that sound azimuth information is redundantly distributed across DCIC top ranked units, which points out that mechanisms beyond coding efficiency could be relevant for this population code.

While the decoding error observed from our DCIC datasets obtained in passively listening, untrained mice coincidentally matches the discrimination ability of highly trained, motivated mice (Lauer et al., 2011), a relationship between decoding error and psychophysical performance remains to be determined. Interestingly, a primary sensory representations should theoretically be even more precise than the behavioral performance as reported in the visual system (Stringer et al., 2021).”

Moreover, the concept of redundancy (of spatial information carried by units throughout the DCIC) is difficult for me to disentangle. One interpretation of this formulation could be that there are non-overlapping populations of neurons distributed across the DCIC that each could predict azimuth independently of each other, which is unlikely what the authors meant. If the authors meant generally that multiple neurons in the DCIC carry sufficient spatial information, then a single neuron would have been able to predict sound source azimuth, which was not the case. I have the feeling that they actually mean "complimentary", but I leave it to the authors to clarify my confusion, should they wish.

We observed that the response patterns from relatively small fractions of the azimuth sensitive DCIC units (4-7 top ranked units) are sufficient to generate an effective code for sound azimuth, while 32-40% of all simultaneously recorded DCIC units are azimuth sensitive. In light of this observation, we interpreted that the azimuth information carried by the population should be redundantly distributed across the complete subpopulation of azimuth sensitive DCIC units.

In summary, the present study represents a significant body of work that contributes substantially to the field of spatial auditory coding and systems neuroscience. However, limitations of the imaging dataset and model as applied in the study muddles concrete conclusions about how the DCIC precisely encodes sound source azimuth and even more so to sound localisation in a behaving animal. Nevertheless, it presents a novel and unique dataset, which, regardless of secondary interpretation, corroborates the general notion that auditory space is encoded in an extraordinarily complex manner in the mammalian brain.

Reviewer #3 (Public Review):

Summary:

Boffi and colleagues sought to quantify the single-trial, azimuthal information in the dorsal cortex of the inferior colliculus (DCIC), a relatively understudied subnucleus of the auditory midbrain. They used two complementary recording methods while mice passively listened to sounds at different locations: a large volume but slow sampling calcium-imaging method, and a smaller volume but temporally precise electrophysiology method. They found that neurons in the DCIC were variable in their activity, unreliably responding to sound presentation and responding during inter-sound intervals. Boffi and colleagues used a naïve Bayesian decoder to determine if the DCIC population encoded sound location on a single trial. The decoder failed to classify sound location better than chance when using the raw single-trial population response but performed

significantly better than chance when using intermediate principal components of the population response. In line with this, when the most azimuth dependent neurons were used to decode azimuthal position, the decoder performed equivalently to the azimuthal localization abilities of mice. The top azimuthal units were not clustered in the DCIC, possessed a contralateral bias in response, and were correlated in their variability (e.g., positive noise correlations). Interestingly, when these noise correlations were perturbed by inter-trial shuffling decoding performance decreased. Although Boffi and colleagues display that azimuthal information can be extracted from DCIC responses, it remains unclear to what degree this information is used and what role noise correlations play in azimuthal encoding.

Strengths:

The authors should be commended for collection of this dataset. When done in isolation (which is typical), calcium imaging and linear array recordings have intrinsic weaknesses. However, those weaknesses are alleviated when done in conjunction with one another - especially when the data largely recapitulates the findings of the other recording methodology. In addition to the video of the head during the calcium imaging, this data set is extremely rich and will be of use to those interested in the information available in the DCIC, an understudied but likely important subnucleus in the auditory midbrain.

The DCIC neural responses are complex; the units unreliably respond to sound onset, and at the very least respond to some unknown input or internal state (e.g., large inter-sound interval responses). The authors do a decent job in wrangling these complex responses: using interpretable decoders to extract information available from population responses.

Weaknesses:

The authors observe that neurons with the most azimuthal sensitivity within the DCIC are positively correlated, but they use a Naïve Bayesian decoder which assume independence between units. Although this is a bit strange given their observation that some of the recorded units are correlated, it is unlikely to be a critical flaw. At one point the authors reduce the dimensionality of their data through PCA and use the loadings onto these components in their decoder. PCA incorporates the correlational structure when finding the principal components and constrains these components to be orthogonal and uncorrelated. This should alleviate some of the concern regarding the use of the naïve Bayesian decoder because the projections onto the different components are independent. Nevertheless, the decoding results are a bit strange, likely because there is not much linearly decodable azimuth information in the DCIC responses. Raw population responses failed to provide sufficient information concerning azimuth for the decoder to perform better than chance. Additionally, it only performed better than chance when certain principal components or top ranked units contributed to the decoder but not as more components or units were added. So, although there does appear to be some azimuthal information in the recoded DCIC populations - it is somewhat difficult to extract and likely not an 'effective' encoding of sound localization as their title suggests.

As described in the responses to reviewers 1 and 2, we chose the naïve Bayes classifier as a decoder to determine the influence of noise correlations through the most conservative approach possible, as this classifier would be least likely to find a contribution of correlated noise. Also, we chose this decoder due to its robustness against limited numbers of trials collected, in comparison to “data hungry” non linear classifiers like KNN or artificial neuronal nets. Lastly, we observed that small populations of noisy, unreliable (do not respond in every trial) DCIC neurons can encode stimulus azimuth in passively listening mice matching the discrimination error of trained mice. Therefore, while this encoding is definitely not efficient, it can still be considered effective.

Although this is quite a worthwhile dataset, the authors present relatively little about the characteristics of the units they've recorded. This may be due to the high variance in responses seen in their population. Nevertheless, the authors note that units do not respond on every trial but do not report what percent of trials that fail to evoke a response. Is it that neurons are noisy because they do not respond on every trial or is it also that when they do respond they have variable response distributions? It would be nice to gain some insight into the heterogeneity of the responses.

The limited number of azimuth trial repetitions that we could collect precluded us from making any quantification of the unreliability (failures to respond) and variability in the response distributions from the units we recorded, as we feared they could be misleading. In qualitative terms, “due to the high variance in responses seen” in the recordings and the limited trial sampling, it is hard to make any generalization. In consequence we referred to the observed response variance altogether as neuronal noise. Considering these points, our datasets are publicly available for exploration of the response characteristics.

Additionally, is there any clustering at all in response profiles or is each neuron they recorded in the DCIC unique?

We attempted to qualitatively visualize response clustering using dimensionality reduction, observing different degrees of clustering or lack thereof across the azimuth classes in the datasets collected from different mice. It is likely that the limited number of azimuth trials we could collect and the high response variance contribute to an inconsistent response clustering across datasets.

They also only report the noise correlations for their top ranked units, but it is possible that the noise correlations in the rest of the population are different.

For this study, since our aim was to interrogate the influence of noise correlations on stimulus azimuth encoding by DCIC populations, we focused on the noise correlations from the top ranked unit subpopulation, which likely carry the bulk of the sound location information. Noise correlations can be defined as correlation in the trial to trial response variation of neurons. In this respect, it is hard to ascertain if the rest of the population, that is not in the top rank unit percentage, are really responding and showing response variation to evaluate this correlation, or are simply not responding at all and show unrelated activity altogether. This makes observations about noise correlations from “the rest of the population” potentially hard to interpret.

It would also be worth digging into the noise correlations more - are units positively correlated because they respond together (e.g., if unit x responds on trial 1 so does unit y) or are they also modulated around their mean rates on similar trials (e.g., unit x and y respond and both are responding more than their mean response rate). A large portion of trial with no response can occlude noise correlations. More transparency around the response properties of these populations would be welcome.

Due to the limited number of azimuth trial repetitions collected, to evaluate noise correlations we used the non parametric Kendall tau correlation coefficient which is a measure of pairwise rank correlation or ordinal association in the responses to each azimuth. Positive rank correlation would represent neurons more likely responding together. Evaluating response modulation “around their mean rates on similar trials” would require assumptions about the response distributions, which we avoided due to the potential biases associated with limited sample sizes.

It is largely unclear what the DCIC is encoding. Although the authors are interested in azimuth, sound location seems to be only a small part of DCIC responses. The authors report responses during inter-sound interval and unreliable sound-evoked responses. Although they have video of the head during recording, we only see a correlation to snout and ear movements (which are peculiar since in the example shown it seems the head movements predict the sound presentation). Additional correlates could be eye movements or pupil size. Eye movement are of particular interest due to their known interaction with IC responses - especially if the DCIC encodes sound location in relation to eye position instead of head position (though much of eye-position-IC work was done in primates and not rodent). Alternatively, much of the population may only encode sound location if an animal is engaged in a localization task. Ideally, the authors could perform more substantive analyses to determine if this population is truly noisy or if the DCIC is integrating un-analyzed signals.

We unsuccessfully attempted eye tracking and pupillometry in our videos. We suspect that the reason behind this is a generally overly dilated pupil due to the low visible light illumination conditions we used which were necessary to protect the PMT of our custom scope.

It is likely that DCIC population activity is integrating un-analyzed signals, like the signal associated with spontaneous behaviors including face movements (Stringer et al., 2019), which we observed at the level of spontaneous snout movements. However investigating if and how these signals are integrated to stimulus azimuth coding requires extensive behavioral testing and experimentation which is out of the scope of this study. For the purpose of our study, we referred to trial-to-trial response variation as neuronal noise. We note that this definition of neuronal noise can, and likely does, include an influence from un-analyzed signals like the ones from spontaneous behaviors.

Although this critique is ubiquitous among decoding papers in the absence of behavioral or causal perturbations, it is unclear what - if any - role the decoded information may play in neuronal computations. The interpretation of the decoder means that there is some extractable information concerning sound azimuth - but not if it is functional. This information may just be epiphenomenal, leaking in from inputs, and not used in computation or relayed to downstream structures. This should be kept in mind when the authors suggest their findings implicate the DCIC functionally in sound localization.

Our study builds upon previous reports by other independent groups relying on “causal and behavioral perturbations” and implicating DCIC in sound location learning induced experience dependent plasticity (Bajo et al., 2019, 2010; Bajo and King, 2012), which altogether argues in favor of DCIC functionality in sound localization.

Nevertheless, we clarified in the discussion of the revised manuscript that a relationship between the observed decoding error and the psychophysical performance, or the ability of the DCIC network to perform Bayesian decoding computations, both remain to be determined (please see responses to Reviewer #2).

It is unclear why positive noise correlations amongst similarly tuned neurons would improve decoding. A toy model exploring how positive noise correlations in conjunction with unreliable units that inconsistently respond may anchor these findings in an interpretable way. It seems plausible that inconsistent responses would benefit from strong noise correlations, simply by units responding together. This would predict that shuffling would impair performance because you would then be sampling from trials in which some units respond, and trials in which some units do not respond - and may

predict a bimodal performance distribution in which some trials decode well (when the units respond) and poor performance (when the units do not respond).

In samples with more than 2 dimensions, the relationship between signal and noise correlations is more complex than in two dimensional samples (Montijn et al., 2016) which makes constructing interpretable and simple toy models of this challenging. Montijn et al. (2016) provide a detailed characterization and model describing how the accuracy of a multidimensional population code can improve when including “positive noise correlations amongst similarly tuned neurons”. Unfortunately we could not successfully test their model based on Mahalanobis distances as we could not verify that the recorded DCIC population responses followed a multivariate gaussian distribution, due to the limited azimuth trial repetitions we could sample.

Significance:

Boffi and colleagues set out to parse the azimuthal information available in the DCIC on a single trial. They largely accomplish this goal and are able to extract this information when allowing the units that contain more information about sound location to contribute to their decoding (e.g., through PCA or decoding on top unit activity specifically). The dataset will be of value to those interested in the DCIC and also to anyone interested in the role of noise correlations in population coding. Although this work is first step into parsing the information available in the DCIC, it remains difficult to interpret if/how this azimuthal information is used in localization behaviors of engaged mice.

Recommendations for the authors:

Reviewer #2 (Recommendations For The Authors):

General:

The manuscript is generally well written, but could benefit from a quick proof by a native English speaker (e.g., "the" inferior colliculus is conventionally used with its article). The flow of arguments is also generally easy to follow, but I would kindly ask the authors to consider elaborating or clarifying the following points (including those already mentioned in my public review).

(1) Choice of model:

There are countless ways one can construct a decoder or classifier that can predict a presented sensory stimulus based on a population neuronal response. Given the assumptions of independence as mentioned in my public review, I would ask the authors to explicitly justify their choice of a naïve Bayesian classifier.

A section detailing the logic of classifier choice is now included in the results section at page 10 and the last paragraph of page 18 from the revised version of the manuscript.

(2) Number of imaging repetitions:

For particularly noisy datasets, 14 repetitions is indeed quite few. I reckon this was not the choice of the authors, but rather limited by the inherent experimental conditions. Despite minimisation of required average laser power during the development of s-TeFo imaging, the authors still required almost 200 mW (which is still quite a lot of exposure). Although 14 repetitions for 13 azimuthal locations every 5 s is at face value a relatively short imaging session (~15 min.), at 191 mW, with the desire to image mice multiple times, I could imagine that this is a practical limitation the authors faced (to avoid excessive tissue heating or photodamage, which was assessed in the original Nature

Methods article, but not here). Nevertheless, this logic (or whatever logic they had) should be explained for non-imaging experts in the readership.

This is now addressed in the answers to the public reviews.

(3) Redundancy:

It is honestly unclear to me what the authors mean by this. I don't speculate that they mean there are "redundant" (small) populations of neurons that sufficiently encode azimuth, but I'm actually not certain. If that were the case, I believe this would need further clarification, since redundant representations would be both inconsistent with the general (perhaps surprising) finding that large populations are not required in the DCIC, which is thought to be the case at earlier processing stages.

In the text we are referring to the azimuth information being redundantly distributed across DCIC top ranked units. We do not mention redundant “populations of neurons”.

(4) Correspondence of decoding accuracy with psychometric functions in mice: While this is an interesting coincidental observation, it should not be interpreted that the neuronal detection threshold in the DCIC somehow is somehow responsible its psychometric counterpart (which is an interesting yet exceedingly complex question). Although I do not believe the authors intended to suggest this, I would personally be cautious in the way I describe this correspondence. I mention this because the authors point it out multiple times in the manuscript (whereas I would have just mentioned it once in passing).

This is now clarified in the revised manuscript.

(5) Noisy vs. sparse:

I'm confident that the authors understand the differences between these terms, both in concept (stochastic vs. scattered) and in context (neuronal vs. experimental), but I personally would be cautious in the way I use them in the description of the study. Indeed, auditory neuronal signals are to my knowledge generally thought to be both sparse and noisy, which is in itself interesting, but the study also deals with substantial experimental (recording) noise, and I think it's important for the readership to understand when "noise" refers to the recordings (in particular the imaging data) and to neuronal activity. I mention this specifically because "noisy" appears in the title.

We have clarified this issue at the bottom of page 5 by adding the following sentences to the revised manuscript:

“In this section we used the word “noise” to refer to the sound stimuli used and recording setup background sound levels or recording noise in the acquired signals. To avoid confusion, from now on in the manuscript the word “noise” will be used in the context of neuronal noise, which is the trial-to-trial variation in neuronal responses unrelated to stimuli, unless otherwise noted.”

(6) More details in the Methods:

The Methods section is perhaps the least-well structured part of the present manuscript in my view, and I encourage the authors to carefully go through it and add the following information (in case I somehow missed it).

a. Please also indicate the number of animals used here.

Added.

b. How many sessions were performed on each mouse?

This is already specified in the methods section in page 25:

“mice were imaged a total of 2-11 times (sessions), one to three times a week.”

We added for clarification:

“Datasets here analyzed and reported come from the imaging session in which we observed maximal calcium sensor signal (peak AAV expression) and maximum number of detected units.”

c. For the imaging experiments, was it possible to image the same units from session to session?

This is not possible for sTeFo 2P data due to low spatial resolution which makes precisely matching neuron ROIs across sessions challenging.

d. Could the authors please add more detail to the analyses of the videos (to track facial movements) or provide a reference?

Added citation.

e. The same goes for the selection of subcellular regions of interest that were used as "units."

Added to page 25:

“We used the CaImAn package (Giovannucci et al., 2019) for automatic ROI segmentation through constrained non negative matrix factorization and selected ROIs (Units) showing clear Ca transients consistent with neuronal activity, and IC neuron somatic shape and size (Schofield and Beebe, 2019).”

Specific: In order to maximise the efficiency of my comments and suggestions (as there are no line numbers), my numerated points are organised in sequential order.

(1) Abstract: I wouldn't personally motivate the study with the central nucleus of the IC (i.e. I don't think this is necessary). I think the authors can motivate it simply with the knowledge gaps in spatial coding throughout the auditory system, in which such large data sets such as the ones presented here are of general value.

(2) Page 4: 15-50 kHz "white" noise is incorrect. It should be "band-passed" noise.

Changed.

(3) Supplemental figure 1, panel A: Since the authors could not identify cell bodies unequivocally from their averaged volume timeseries data, it would be clearer to the readership if larger images are shown, so that they can evaluate (speculate) for themselves what subcellular structures were identified as units. Even better would be to include a planar image through a cross-section. As mentioned above, not everything determined for the cortex or hippocampus can be assumed to be true for the DCIC.

The raw images and segmentations are publicly available for detailed inspections.

(4) Supplemental figure 2, panel A: This panel requires further explanation, in particular the panel on the right. I assume that to be a simple subtraction of sequential frames, but I'm thrown off by the "d(Grey)" colour bar. Also, if "grey" refers to the neutral colour, it is conventionally spelled "gray" in US-American English.

Changed.

(5) Supplemental figure 2, panel B: I'm personally curious why the animals exhibited movement just prior to a stimulus. Did they learn to anticipate the presentation of a sound after some habituation? Is that somehow a pre-emptive startle response? We observe that in our own experiments (but as we stochastically vary the inter-trial-intervals, the movement typically occurs directly after the stimulus). I don't suggest the authors dwell on this, but I find it an interesting observation.

It is indeed interesting, but we can't conclude much about it without comparing it to random inter-trial-intervals.

(6) Supplemental figure 3: I personally find these data (decoding of all electrophysiological data) of central relevance to the study, since it mirrors the analyses presented for its imaging data counterpart and encourage the authors to move it to the main text.

Changed.

(7) Page 12: Do the authors have any further analyses of spatial tuning functions? We all know they can parametrically obscure (i.e., bi-lobed, non-monotonic, etc.), but having these parameters (even if just in a supplemental figure) would be informative for the spatial auditory community.

We dedicated significant effort to attempt to parametrize and classify the azimuth response dependency functions from the recorded DCIC cells in an unbiased way. Nevertheless, given the observed response noise and the "obscure" properties of spatial tuning functions mentioned by the reviewer, we could only reach the general qualitative observation of having a more frequent contralateral selectivity.

(8) Page 14 (end): Here, psychometric correspondence is referenced. Please add the Lauer et al., (2011) reference, or, as I would, remove the statement entirely and save it for the discussion (where it is also mentioned and referenced).

Changed.

(9) Figure 5, Panels B and C: Why don't the authors report the Kruskal-Wallis tests (for increasing number of units training the model), akin to e.g., Panel G of Figure 4? I think that would be interesting to see (e.g., if the number of required units to achieve statistical significance is the same).

Within class randomization produced a moderate effect on decoder performance, achieving statistical significance at similar numbers of units, as seen in figure 5 panels B and C. We did not include these plots for the sake of not cluttering the figure with dense distributions and fuzzing the visualization of the differences between the distributions shown.

(10) Figure 5, Panels B and C (histograms): I see a bit of skewedness in the distributions (even after randomisation). Where does this come from? This is just a small

talking point.

We believe this is potentially due to more than one distribution of pairwise correlations combined into one histogram (like in a Gaussian mixture model).

(11) Page 21: Could the authors please specify that the Day and Delgutte (2013) study was performed on rabbits? Since rabbits have an entirely different spectral hearing range compared to mice, spatial coding principles could very well be different in those animals (and I'm fairly certain such a study has not yet been published for mice).

Specified.

(12) Page 22: I'd encourage the authors to remove the reference to Rayleigh's duplex theory, since mice hardly (if at all) use interaural time differences for azimuthal sound localisation, given their generally high-frequency hearing range.

That sentence is meant to discuss beyond the mouse model an exciting outlook of our findings in light of previous reports, which is a hypothetical functional relationship between the tonotopy in DCIC and the spatial distribution of azimuth sensitive DCIC neurons. We have clarified this now in the text.

(13) Page 23: I believe the conventional verb for gene delivery with viruses is still "transduce" (or "infect", but not "induce"). What was the specific "syringe" used for stereotactic injections? Also, why were mice housed separately after surgery? This question pertains to animal welfare.

Changed. The syringe was a 10ml syringe to generate positive or negative pressure, coupled to the glass needle through a silicon tubing via a luer 3-way T valve. Single housing was chosen to avoid mice compromising each other's implantations. Therefore this can be seen as a refinement of our method to maximize the chances of successful imaging per implanted mouse.

(14) Page 25: Could the authors please indicate the refractory period violation time window where? I had to find it buried in the figure caption of Supplementary figure 1.

Added.

(15) Page 27: What version of MATLAB was used? This could be important for reproduction of the analyses, since The Mathworks is infamously known to add (or even more deplorably, modify) functions in particular versions (and not update older ones accordingly).

Added.

Reviewer #3 (Recommendations For The Authors):

Overall I thought this was a nice manuscript and a very interesting dataset. Here are some suggestions and minor corrections:

You may find this work of interest - 'A monotonic code for sound azimuth in primate inferior colliculus' 2003, Groh, Kelly & Underhill.

We thank the reviewer for pointing out this extremely relevant reference, which we regrettably failed to cite. It is now included in the revised version of the manuscript.

In your introduction, you state "our findings point to a functional role of DCIC in sound location coding". Though your results show that there is azimuthal information contained in a subset of DCIC units there's no evidence in the manuscript that shows a functional link between this representation and sound localization.

This is now addressed in the answers to the public reviews.

I found the variability in your DCIC population quite striking - especially during the intersound intervals. The entrainment of the population in the imaging dataset suggests some type of input activating the populations - maybe these are avenues for further probing the variability here:

(1) I'm curious if you can extract eye movements from your video. Work from Jennifer Groh shows that some cells in the primate inferior colliculus are sensitive to different eye positions (Groh et. al., 2001). With recent work showing eye movements in rodents, it may explain some of the variance in the DCIC responses.

This is now addressed in the answers to the public reviews.

(2) I was also curious if the motor that moves the speaker made noise It could be possible some of the 'on going' activity could be some sound-evoked response.

We were careful to set the stepper motor speed so that it produced low frequency noise, within a band mostly outside of the hearing range of mice (<4kHz). Nevertheless, we cannot fully rule out that a very quiet but perhaps very salient component of the motor noise could influence the activity during the inter trial periods. The motor was stationary and quiet for a period of at least one stimulus duration before and during stimulus presentation.

(3) Was the sound you present frozen or randomly generated on each trial? Could there be some type of structure in the noise you presented that sometimes led cells to respond to a particular azimuth location but not others?

The sound presented was frozen noise. This is now clarified in the methods section.

It may be useful to quantify the number of your units that had refractory period violations.

Our manual curation of sorted units was very stringent to avoid mixing differently tuned neurons. The single units analyzed had very infrequent refractory period violations, in less than ~5% of the spikes, considering a 2 ms refractory period.

Was the video recording contralateral or ipsilateral to the recording?

The side of the face ipsilateral to the imaged IC was recorded. Added to methods.

I was struck by the snout and ear movements - in the example shown in Supplementary Figure 2B it appears as they are almost predicting sound onset. Was there any difference in ear movements in the habituated and non-habituated animals? Also, does the placement of the cranial window disturb any of the muscles used in ear movement?

Mouse snout movements appear to be quite active perhaps reflecting arousal (Stringer et al., 2019). We cannot rule out that the cranial window implantation disturbed ear movement but

while moving the mouse headfixed we observed what could be considered normal ear movements.

Did you correlate time-point by time-point in the average population activity and movement or did you try different temporal lags/leads in case the effect of the movements was delayed in some way?

Point by point due to 250ms time resolution of imaging.

Are the video recordings only available during the imaging? It would be nice to see the same type of correlations in the neuropixel-acquired data as well.

Only imaging. For neuropixels recordings, we were skeptical about face videography as we suspected that face movements were likely influenced by the acute nature of the preparation procedure. Our cranial window preparation in the other hand involved a recovery period of at least 4 weeks. Therefore we were inclined to perform videographical interrogation of face movements on these mice instead.

If you left out more than 1 trial do you think this would help your overfitting issue (e.g. leaving out 20% of the data).

Due to the relatively small number of trial repetitions collected, fitting the model with an even smaller training dataset is unlikely to help overfitting and will likely decrease decoder performance.

It would be nice to see a confusion matrix - even though azimuthal error and cumulative distribution of error are a fine way to present the data - a confusion matrix would tell us which actual sounds the decoder is confusing. Just looking at errors could result in some funky things where you reduce the error generally but never actually estimate the correct location.

We considered confusion matrices early on in our study but they were not easily interpretable or insightful, likely due to the relatively low discrimination ability of the mouse model with +/- 30° error after extensive training. Therefore, we reasoned that in passively listening mice (and likely trained mice too) with limited trial repetitions, an undersampled and diffuse confusion matrix is expected which is not an ideal means of visualizing and comparing decoding errors. Hence we relied on cumulative error distributions.

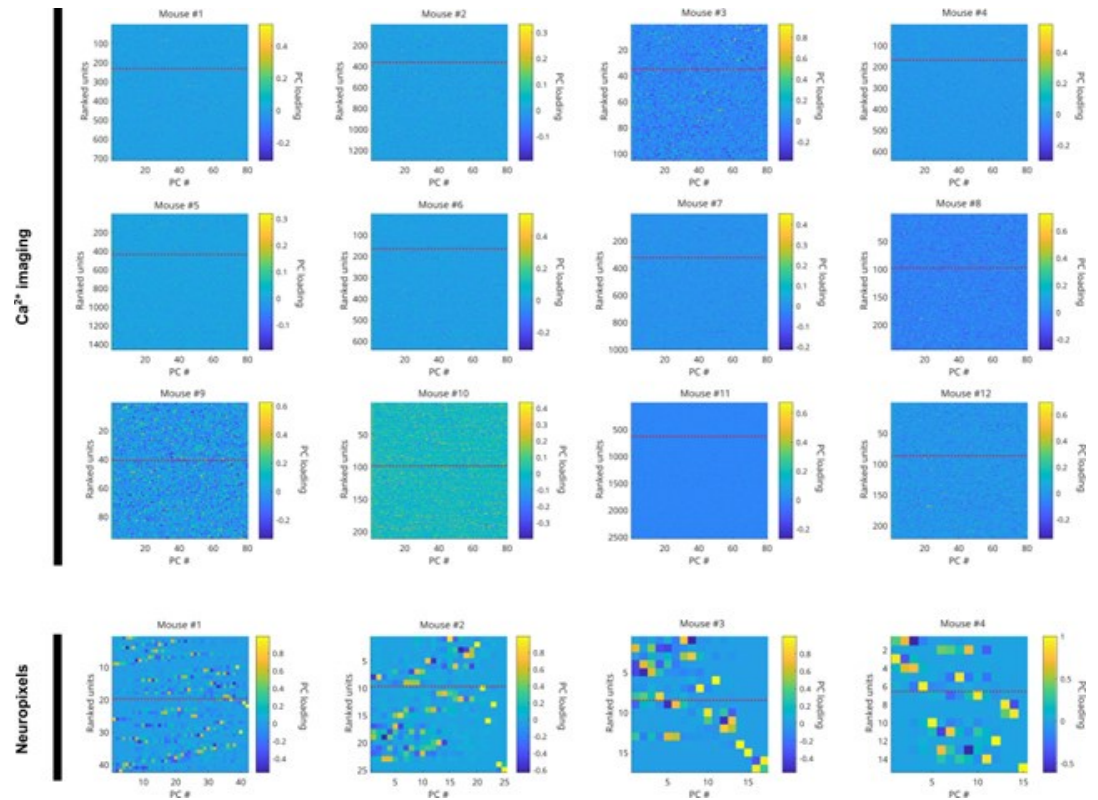
Do your top-ranked units have stronger projections onto your 10-40 principal components?

It would be interesting to know if the components are mostly taking into account those 30ish percent of the population that is dependent upon azimuth.

Inspection of PC loadings across units ranked based on response dependency to stimulus azimuth does not show a consistent stronger projection of top ranked units onto the first 10-40 principal components (Author response image 3).

Author response image 3.

PC loading matrices for each recorded mouse. The units recorded in each mouse are ranked in descending order of response dependency to stimulus azimuth based on the p value of the chi square test. Units above the red dotted line display a chi square p value < 0.05, units below this line have p values >= 0.05.



How much overlap is there in the tuning of the top-ranked units?

This is quite varying from mouse to mouse and imaging vs electrophysiology, which makes it hard to make a generalization since this might depend on the unique DCIC population sampled in each mouse.

I'm not really sure I follow what the nS/N adds - it doesn't really measure tuning but it seems to be introduced to discuss/extract some measure of tuning.

nS/N is used to quantify how noisy neurons are, independent of how sensitive their responses are to the stimulus azimuth.

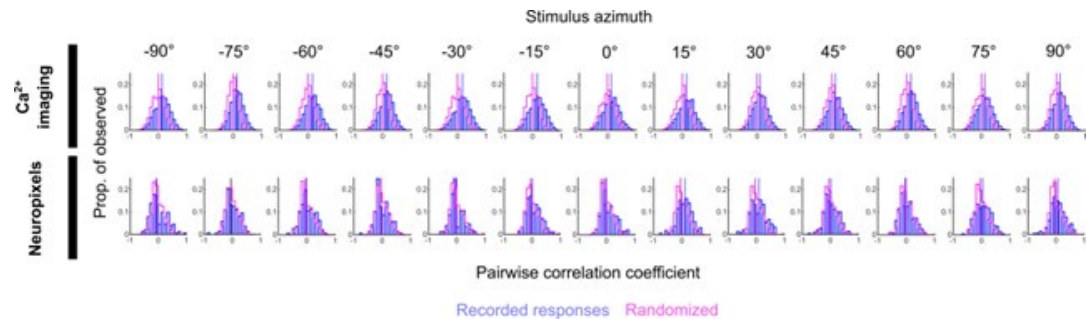
Is the noise correlation - observed to become more positive - for more contralateral stimuli a product of higher firing rates due to a more preferred stimulus presentation or a real effect in the data? Was there any relationship between distance and strength of observed noise correlation in the DCIC?

We observed a consistent and homogeneous trend of pairwise noise correlation distributions either shifted or tailed towards more positive values across stimulus azimuths, for imaging and electrophysiology datasets (Author response image 3). The lower firing frequency observed in neuropixels recordings in response to ipsilateral azimuths could have affected the statistical power of the comparison between the pairwise noise correlation coefficient distribution to its randomized chance level, but the overall histogram shapes qualitatively support this consistent trend across azimuths (Author response image 4).

Author response image 4.

Distribution histograms for the pairwise correlation coefficients (Kendall tau) from pairs of simultaneously recorded top ranked units across mice (blue) compared to the chance level

distribution obtained through randomization of the temporal structure of each unit's activity to break correlations (purple). Vertical lines show the medians of these distributions. Imaging data comes from $n = 12$ mice and neuropixels data comes from $n = 4$ mice.



Typos:

'a population code consisting on the simultaneous' > should on be of?

'half of the trails' > trails should be trials?

'referncing the demuxed channels' > should it be demixed?

Corrected.

<https://doi.org/10.7554/eLife.97598.2.sa0>