Report

Current Biology

Source identity shapes spatial preference in primary auditory cortex during active navigation

Highlights

- Localization task during free navigation prompts diverse spatial tuning in gerbil A1
- Spatial preference of individual neurons changes with sound-source identity
- Ego- and allocentric information are spatiotemporally coexistent in A1 ensembles
- Active sensing reveals new cortical representations for sensory object identification

Authors

Diana Amaro, Dardo N. Ferreiro, Benedikt Grothe, Michael Pecka

Correspondence

pecka@bio.lmu.de

In brief

Using a new localization paradigm that features self-motion and distinct relevance of individual sound sources, Amaro et al. show that the spatial tuning in primary auditory cortex is highly diverse and not—as canonically assumed—purely egocentric. Strikingly, the tuning of most neurons is altered by the task-specific identity of the source.

Amaro et al., 2021, Current Biology *31*, 3875–3883
September 13, 2021
https://doi.org/10.1016/j.cub.2021.06.025

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Source identity shapes spatial preference in primary auditory cortex during active navigation

Diana Amaro,[1,2] Dardo N. Ferreiro,[1,3] Benedikt Grothe,[1,2,4] and Michael Pecka[1,5,*]
[1]Division of Neurobiology, Department Biology II, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany
[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany
[3]Department of General Psychology and Education, Ludwig-Maximilians-Universität München, Germany
[4]Max Planck Institute of Neurobiology, Planegg-Martinsried, Germany
[5]Lead contact
*Correspondence: pecka@bio.lmu.de
https://doi.org/10.1016/j.cub.2021.06.025

SUMMARY

Information about the position of sensory objects and identifying their concurrent behavioral relevance is vital to navigate the environment. In the auditory system, spatial information is computed in the brain based on the position of the sound source relative to the observer and thus assumed to be egocentric throughout the auditory pathway. This assumption is largely based on studies conducted in either anesthetized or head-fixed and passively listening animals, thus lacking self-motion and selective listening. Yet these factors are fundamental components of natural sensing[1] that may crucially impact the nature of spatial coding and sensory object representation.[2] How individual objects are neuronally represented during unrestricted self-motion and active sensing remains mostly unexplored. Here, we trained gerbils on a behavioral foraging paradigm that required localization and identification of sound sources during free navigation. Chronic tetrode recordings in primary auditory cortex during task performance revealed previously unreported sensory object representations. Strikingly, the egocentric angle preference of the majority of spatially sensitive neurons changed significantly depending on the task-specific identity (outcome association) of the sound source. Spatial tuning also exhibited large temporal complexity. Moreover, we encountered egocentrically untuned neurons whose response magnitude differed between source identities. Using a neural network decoder, we show that, together, these neuronal response ensembles provide spatiotemporally co-existent information about both the egocentric location and the identity of individual sensory objects during self-motion, revealing a novel cortical computation principle for naturalistic sensing.

RESULTS AND DISCUSSION

Localizing and identifying sensory objects while navigating the environment are fundamental brain functions. Auditory neurons compute azimuthal spatial information via angle-specific differences of sound features between the ears. From brainstem to primary auditory cortex (A1), a predominance of broad neuronal tuning to contralateral sound-source locations has been reported with a smaller subset tuned to ipsilateral or frontal positions.[3–7] Accordingly, auditory neuronal spatial tuning is canonically regarded to be egocentric. Notably, while self-movement constantly alters the egocentric sound-source location, the perception of source position remains stable relative to the world coordinates, i.e., is allocentric.[8] Recent results from one study with freely moving ferrets suggest the existence of purely allocentric representation in a small minority of neurons in A1.[9] However, the stimuli and associated sources in this study were task-irrelevant since subjects were passively exposed to sounds while searching for water. Yet active sensing and task engagement/stimulus relevance critically modulate neuronal coding in A1,[10–16] and thus it remains unclear how sound sources are represented during natural sensing behavior. Here, we took advantage of the recently developed Sensory Island Task (SIT) paradigm[17] and recorded from A1 neurons of freely exploring animals that actively localized sounds from sources with distinct task identities (i.e., associated behavioral outcome) and allocentric locations (Figures 1A–1E). We found that the spatial tuning in A1 during active sensing deviated profoundly from the canonically assumed egocentric representation. The majority of neurons exhibited temporally diverse spatial tuning that differed between sound sources. Artificial neural network decoding demonstrated that on the population level these novel tuning features generate spatiotemporally coexistent information about the instantaneous source angles and angle-independent source identity.

Gerbils identify reward-associated sound source based on spatial cues

To investigate potentially undiscovered spatial representations during active sensing, we trained Mongolian gerbils (*Meriones unguiculatus*) to report whenever the presentation of a pulsed harmonic stack switched from a "background" loudspeaker to a "target" loudspeaker (separated by 180° in a circular arena). The sound-source change was triggered whenever the gerbil entered a specific target area in the arena (the "island"),

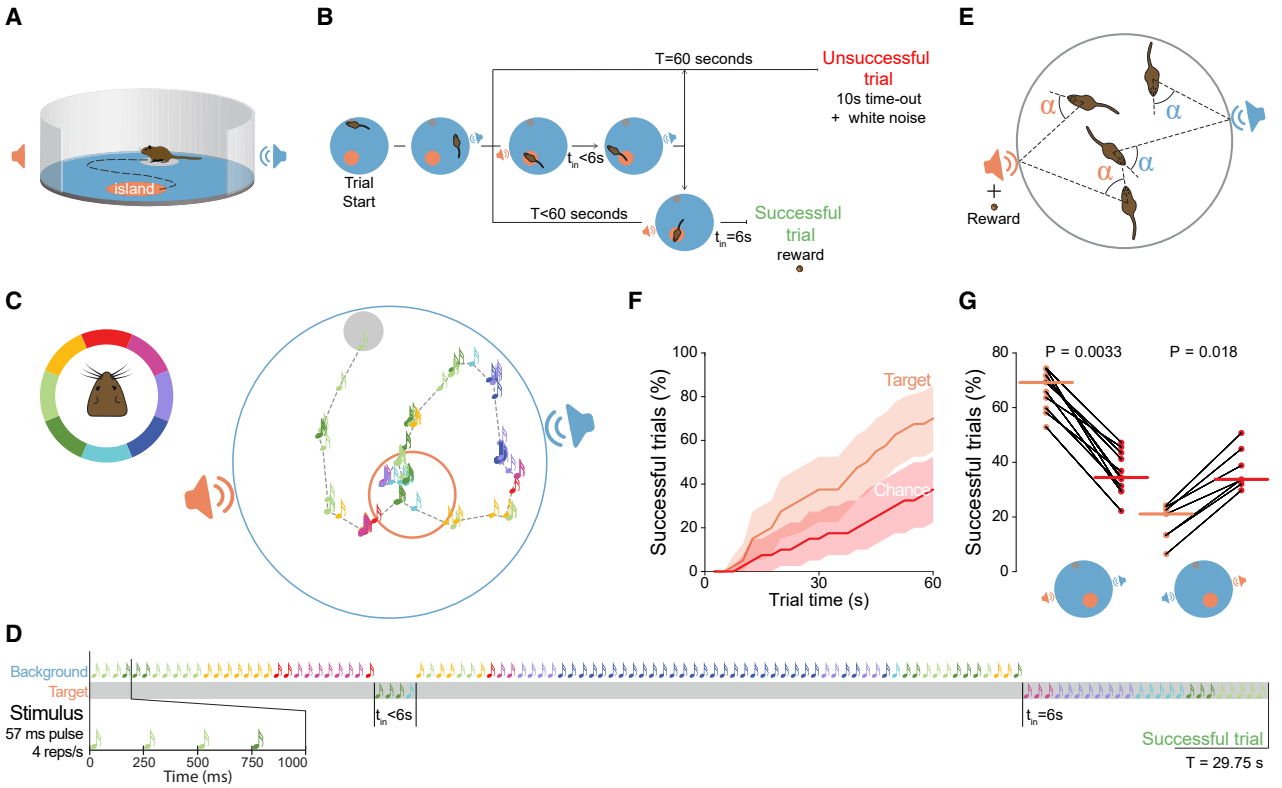

Figure 1. Sound localization task

(A–E) Scheme of the SIT setup and paradigm. See also Video S1.

(A) A freely moving gerbil is inside a circular arena with a slightly elevated initiation platform and two loudspeakers located diametrically opposed to each other. A randomly located island is associated to sound played from the target loudspeaker (orange) and all other locations are associated to sound played from the background loudspeaker (blue).

(B) The animal starts a trial in the initiation platform and hears a pulsed harmonic complex (4 Hz repetition rate) played from the background loudspeaker until it enters the island, which triggers a switch to the target loudspeaker; to report the switch the animal must stay within the island for 6 s (SIT-time). In that case, the trial is considered successful and a food reward is provided to the animal. If within 60 s of the trial start the animal does not find the island or does not remain inside it for 6 s continuously (i.e., exited prematurely), the trial is considered unsuccessful and white noise is presented from the target loudspeaker together with a 10 s time-out during which the animal cannot start another trial.

(C) By exploring the arena in search of the island, the egocentric sound-source location is modulated. The left panel illustrates 8 color-coded bins of the azimuthal position of the active loudspeaker relative to the animal and the right panel shows the trajectory of the animal during a successful trial, in which colored note symbols indicate the position of the animal at the moment of each sound presentation and corresponding egocentric sound-source location (color-coded as represented in the left panel).

(D) Temporal representation of the trajectory depicted in (C). Each note represents a 57 ms harmonic complex that was played every 250 ms either by the background loudspeaker (upper row) or by the target loudspeaker (gray lower row), and its color represents the egocentric sound-source location color-coded as depicted in (C).

(E) Egocentric sound-source angles are ambiguous (as the same angle can be produced by either source) and hence insufficient for task completion. Association of the target loudspeaker and its allocentric location with reward delivery allows the formation of source-specific identities.

(F) Percentage of successful trials (orange) of one animal in comparison to chance (red, STAR Methods) during a particular session as a function of time spent in the trial (40 trials; shaded areas correspond to 95% confidence interval).

(G) Left: percentage of successful trials (orange) in comparison to chance (red) in trials with typical reward contingency (11 animals, Wilcoxon signed-rank test, horizontal lines represent the median, previously reported[17]). Right: percentage of successful trials (orange) in comparison to chance (red) in catch trials with reversal of the reward contingency (5 animals, Wilcoxon signed-rank test, horizontal lines represent the median, previously reported[17]).

See also Figure S1.

whose position was randomized across trials (Figures 1A and 1B; note that the term "background" does not imply a constant, unattended stimulus, but the "reference" compared to the "target"). The animals were trained to report the detection of activity from the target loudspeaker by remaining within the island for 6 s ("SIT-time") to receive a food reward (Figures 1B–1D; Video S1). Thus, activity of the target loudspeaker was associated with reward delivery. The sounds emitted by the two loudspeakers were spectrally identical (Figure S1A; see below).

Expert animals reliably exhibited highly significant performance levels in reporting the activity of the target loudspeaker (Figures 1F and 1G; evidenced by both the non-overlapping confidence intervals between target and chance level as well as by the consistent higher percentage of successful trials on the target compared to chance). The time spent in the island

(Figure S1C), the egocentric angle upon island entry (Figure S1D), and a hemispheric sound-source separation (Figure S1E) influenced performance, suggesting that accumulation of egocentric spatial information within each trial was critical for detecting activity of the target loudspeaker. Since sounds were limited to low frequencies and localization had to be performed in the horizontal plane, we assume that the predominant egocentric cues were interaural time differences.

Importantly, because the egocentric sound-source location can be the same for both loudspeakers, this information by itself was rendered insufficient for solving the task (Figure 1E). The animals' performances (Figure 1G, left panel) thus demonstrate that allocentric identification contributed to the task. The formation of such a world-based representation was readily available by referring to visual cues (e.g., "loudspeaker to the west of the initiation platform"). Note that switching the rewarded loudspeaker identity in catch trials resulted in performances significantly below chance level (Figure 1G, right panel), corroborating a task-specific identification of the allocentric location of loudspeakers by the animals and refuting an unspecific change-detection strategy. To test if the animals were sensitive to potentially existing loudspeaker-specific spectral cues to solve the task, we physically swapped the loudspeakers (but not their task-specific identity). Performance remained similar before (mean success rate above chance level ± 95% confidence interval: 19.7% ± 5.7%) and after (20.2% ± 5.1%) swapping (Figure S1B); hence, successful trial completion was achieved by determining the location of the active sound source.

Together, these data demonstrate that the gerbils performed active localization during unrestricted exploration using both egocentric and allocentric spatial information. This allowed us to study how listening to two sound sources with distinct behavioral meaning that are defined by their world-based position influences the spatial representation in A1 during self-motion.

Diverse spatial tuning in A1 during active localization

To investigate the neuronal processing during task performance, expert animals were implanted with custom-made tetrodes to allow chronic recordings of action potentials from neurons in A1 (Figure 2A). Overall, we acquired responses from 364 single neurons and 246 multi-units from 5 gerbils. Activity in A1 was strongly correlated with the subjects' goal-specific behavior, as neuronal firing rates to the target loudspeaker were significantly lower on successful instances compared to those in which animals wrongly left the island before reaching SIT-time ($p = 4 \times 10^{-19}$, Wilcoxon signed-rank test; Figures 2B and S2A–S2C).

Neuronal firing rates exhibited significant modulation (see below and STAR Methods) as a function of the active loudspeaker's angle relative to the animal's body axis. Figure 2C illustrates the activity of two example neurons (recorded simultaneously in one session) over multiple trials as a function of the egocentric angle to the active sound source. The two neurons exhibited marked differences in the preferred angles that triggered action potential firing (evidenced by the color differences).

To determine the spatial tuning from all recorded neurons, we grouped neuronal responses into three response periods: (1) the "onset period," ranging from the determined neuronal response latency (STAR Methods) + stimulus duration of 57 ms (118 neurons exhibited significant egocentric spatial tuning in this response period); (2) the "offset period," offset latency or end of onset period + 57 ms (STAR Methods; 58 neurons); and (3) the "late response period," remaining time between the periodic response to the stimuli (42 neurons; compare Figure 1D).

Construction of polar plots from neuronal responses displayed a large variety of egocentric spatial sensitivity during the onset period (Figure 2D). Alongside a fraction of "canonical" neurons that were similarly tuned to either loudspeaker with contralateral or ipsilateral preference ("i" and "iii" in Figure 2D; onset period, 37/118 = 31.3%; offset period, 7/58; late response period, 0/42), we observed neurons whose tuning to either loudspeaker can be described as orientation-sensitive (onset period, 16/118 = 13.6%; offset period, 1/58; late period, 2/42), with an apparent bias to the front/back orientation ("ii" in Figure 2D; comparable to reports in awake primates[19]).

Identity-specific egocentric tuning

Remarkably, the spatial tuning of a large fraction of the neurons was source-sensitive, as their responses differed between the two loudspeakers (Figures 2D, 3A, and S2D). These neurons were either only significantly spatially tuned to one of the two sound sources ("iv" in Figures 2D and 3A; 47/65 = 72.3% during the onset period) or exhibited a large tuning difference between the two sources (difference in preferred egocentric angle >90°; "v" in Figures 2D and 3A; 18/65 = 27.7% during the onset period; STAR Methods). This variety in spatial tuning types was evident during all three response periods inclusively during the onset period (and in cells with short onset latencies; Figure S2E), and particularly prominent during the late response period (onset period, 65/118 = 55.1%; offset response, 46/58 = 79.3%; late response period, 41/42 = 97.6%; Figure S2D). Source-specific tuning also occurred on the level of multi-unit responses (Figure S2F). Quantification of spatial selectivity exposed population-wide source-specific tuning, since the vector strength (STAR Methods) was significantly higher for the target loudspeaker (Figure 3B).

A dependency of spatial tuning preferences on the putative neuronal type was not observed (regular or fast spiking; Figures 3A and S3A–S3E). Nonetheless, putative regular spiking neurons exhibited sharper spatial tuning to either sound source (Figures S3F and S3G).

Notably, in a sizable fraction of neurons that were classified as egocentrically untuned during all the response periods to both loudspeakers, the firing rate magnitude differed significantly between the two loudspeakers (16/40, 40%; Figure 3A; "vi" in Figure 2D). Since the two loudspeakers' identities were defined by their allocentric location, these neurons could be classified as "purely" allocentric, i.e., sound-source angle-independent coding.

These data suggest that active localization and unrestricted exploration in SIT reveal a variety of previously unreported characteristics of spatial sensitivity. However, movement can modulate cortical activity in a task-unrelated manner.[20,21] We therefore performed additional control analyses with data restricted to neuronal responses from periods when the animals were not moving (no translational or rotatory body movements). This resulted in qualitatively similar observations (Figure S2G), demonstrating that our findings are not caused by unspecific movement modulation.

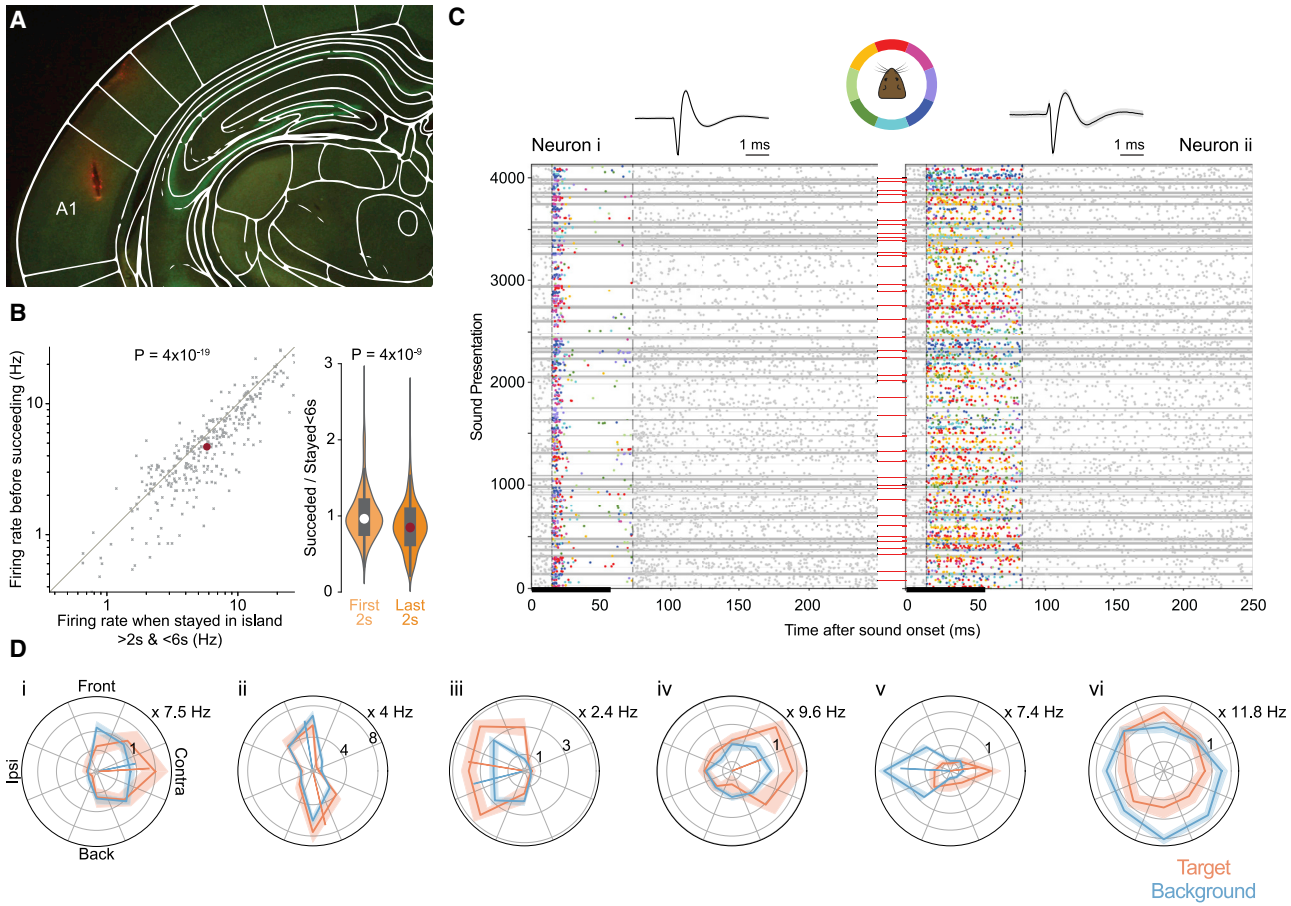

Figure 2. Non-canonical spatial tuning

(A) Reconstruction of tetrode position from histological information, superimposed with the location of A1.[18]

(B) Left: firing rates were significantly lower in the last 2 s before reaching SIT-time in successful trials compared to the last 2 s of premature exits before reaching SIT-time (only instances in which the animal had remained inside for at least 2 s were considered). Each cross represents the mean firing rates of one neuron and the red dot corresponds to the median of the neuronal responses (N = 321 neurons, Wilcoxon signed-rank test, log-log scale). Right: the difference in firing rate between successful trials and premature exits was enhanced in later times (last 2 s) relative to island entry (first 2 s). Violin plots surrounding boxplots (the median is represented by a large dot) depict the distribution of the ratios between the mean firing rate of each neuron during either the first or the last 2 s in situations when the animal successfully finished a trial and when it wrongly left the island (Mann-Whitney U test; first 2 s, N = 346 neurons; last 2 s, N = 321 neurons; same data as shown in the left panel).

(C) Raster plots of neuronal activity in two A1 neurons during task performance in one full session (same as in Figure 1F). The relative timing of each spike within the respective 250 ms period of each single sound presentation in this session is plotted on the x axis and all sound presentations are stacked vertically (y axis). Every spike that occurred during the onset period is color-coded with the egocentric sound-source location at the moment of the sound presentation that preceded the spike as depicted in the color scheme (top). The first vertical line in each plot represents the latency of the respective neuron and the second vertical line represents the end of the onset response period (STAR Methods). Periods of stimulation by the target loudspeaker are highlighted by gray areas. The solid black horizontal lines at the bottom of each plot represent the sound stimulation period. The red horizontal lines between the two plots represent the beginning of new trials. The waveform of each neuron is represented on top of the respective raster plot (solid line, median; shaded area, standard deviation).

(D) Spatial tuning during the onset period for six representative neurons for the target (orange) and background (blue) loudspeakers. The solid line corresponds to the median and the shaded area to the 95% confidence interval. The colored straight lines indicate the preferred egocentric angle for significantly spatially tuned neurons (during the onset period; STAR Methods), with the length scaling with the vector strength. Neurons i and ii correspond to the neurons depicted in (C). See also Figure S2.

Neuronal response timing relative to stimulus duration is known to provide additional information about sound-source location.[4,22,23] We found that the spatial representations in A1 of the actively localizing gerbils also depended on the relative timing of responses, as the egocentric tuning of individual neurons frequently varied considerably across response periods (Figures 3C and 3D). Additionally, this change of preferred egocentric location across response periods occurred typically only for one loudspeaker.

Interlaced population coding of egocentric and allocentric information

To test to what extent the diverse activity patterns that we observed may facilitate distinguishing the two sound sources during task performance, we implemented an artificial neural network model with one hidden layer (multilayer perceptron classifier; STAR Methods). The algorithm was trained on the action potential responses (compare Figure 2C) from 224 units (141 single neurons and 83 multi-units) from two animals from a total of

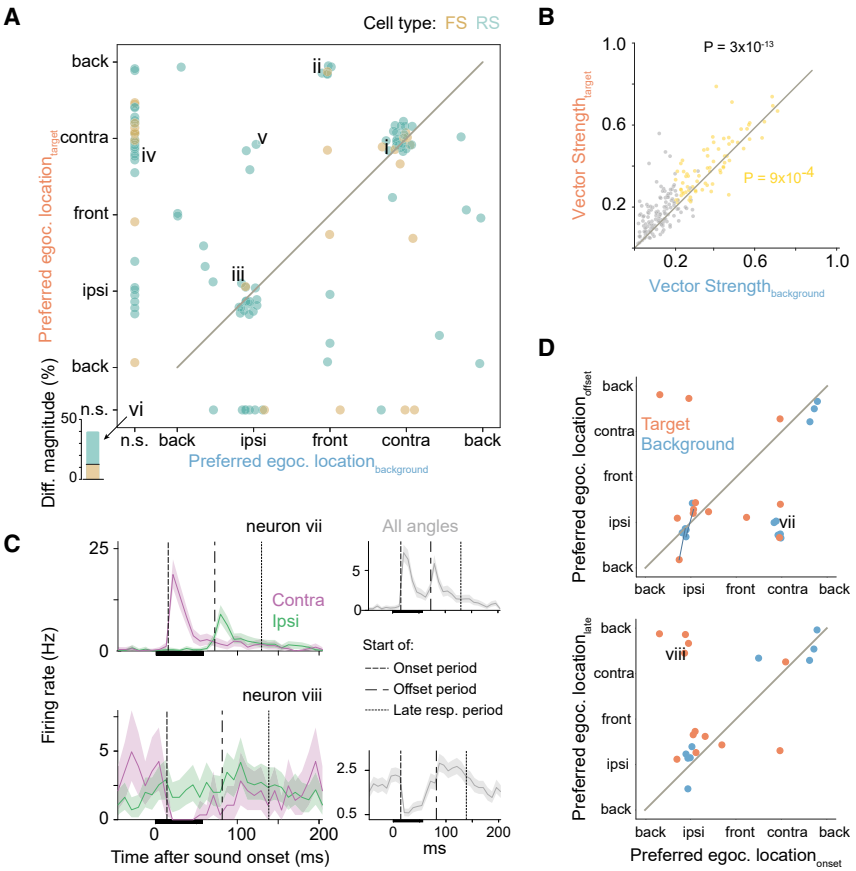

Figure 3. Complexity of neuronal responses in a context-dependent manner

(A) Correspondence of preferred egocentric sound-source locations to the target and to the background loudspeakers calculated for neuronal responses during the onset period. Color code depicts putative neuronal types: FS, fast spiking; RS, regular spiking (118 neurons, 71 of which tuned for both loudspeakers). The roman numerals represent the neurons whose spatial tuning was depicted in Figure 2D. The neurons on the gray diagonal line correspond to those whose preferred egocentric sound-source location does not depend on the identity of the loudspeaker. Inset in lower left corner shows percentage of neurons that were not significantly (n.s.) tuned to either loudspeaker, but exhibited a significant difference in overall response magnitude (16/40 neurons, 5 FS, 11 RS, 8 larger to the target loudspeaker, 8 larger to the background loudspeaker).

(B) Vector strength was larger for the neuronal responses to the target than to the background loudspeaker (Wilcoxon signed-rank test). Yellow data points depict neurons spatially tuned to both loudspeakers (71 neurons), and data from neurons that were not significantly spatially modulated for at least one of the loudspeakers are shown in gray (133 neurons).

(C) Peristimulus time histograms (PSTHs) for two representative neurons with temporally complex spatial tunings. The solid line corresponds to the median, the shaded area to the 95% confidence interval, the solid black horizontal lines at the bottom of each plot represent the sound stimulation period, and the vertical lines indicate the beginning of all the time periods used for the construction of the spatial tuning (i.e., onset, offset, and late response periods). The violet/green PSTHs on the left panel correspond to responses acquired only for contra- or ipsilateral loudspeaker locations (in a 90° angular bin). The gray PSTHs on the right panel disregard the loudspeaker location (i.e., includes all possible angles) and correspond to the same neurons presented on the left panel.

(D) Preferred egocentric sound-source location comparison between the onset and the offset time periods (upper panel, 23 neurons) and between the onset and the late response periods (lower panel, 21 neurons). The dark blue lines connect background and target preferred egocentric sound-source locations that correspond to the same neuron. The roman numerals represent the neurons whose PSTHs were depicted in (C).

See also Figures S2 and S3.

21 sessions, and the training classes were a combination of egocentric spatial information and allocentric sound-source identity (using random under-sampling to prevent class imbalance), in which the 360° around the animal were divided into 8 classes for each loudspeaker (16 classes in total). The neural network was then presented with test data for decoding of specific spatial information. Initially, the data from both loudspeakers were merged to calculate the accuracy of the decoder on the egocentric location of the active loudspeaker. The neural network model classified the egocentric locations with high accuracy (54.2% ± 1.7%) compared to chance levels (12.5%; Figure 4A). Next, we determined the capability of the algorithm in identifying the active loudspeaker irrespective of the egocentric location. Remarkably, the decoding accuracy for identifying the sound source (82.5% ± 1.7%) was also highly increased relative to chance level (50%; Figure 4B). Together, these results demonstrate the coexistence of both subject-based and world-based reference frames in the neuronal representations across the analyzed ensembles (Figure 4C). To better understand the dynamics of the encoded components, a temporal analysis of the decoded information was performed based on the population firing rates across the 250 ms after the onset of each pulse. This analysis indicates that egocentric positions were largely encoded within the first 100 ms after onset of each sound presentation (Figure 4D, inset), whereas the information about the identity of the loudspeaker increased monotonically even after cessation of a given sound presentation (Figure 4D) and reached behavioral performance levels (Figure 1G).

The contribution of the diverse response patterns across the population of neurons to the egocentric and identity decoding accuracy was explored by selectively eliminating units with specific tuning characteristics in the test data and re-assessing the decoding accuracy (Figure 4E; note the marked difference of achieved accuracy compared to chance level, indicated by dashed lines). Elimination of either all units with differential spatial tuning for the two sound sources (18 units; compare Figure 2D, type "v") or exclusion of the same number of units with either diverse response profiles (randomly selected) or exclusively canonic hemispherically tuned units (compare Figure 2D, types "i" and "iii") resulted in only mild reductions of accuracy for egocentric location and identity. Overall, the neural network algorithm demonstrated remarkable robustness toward these

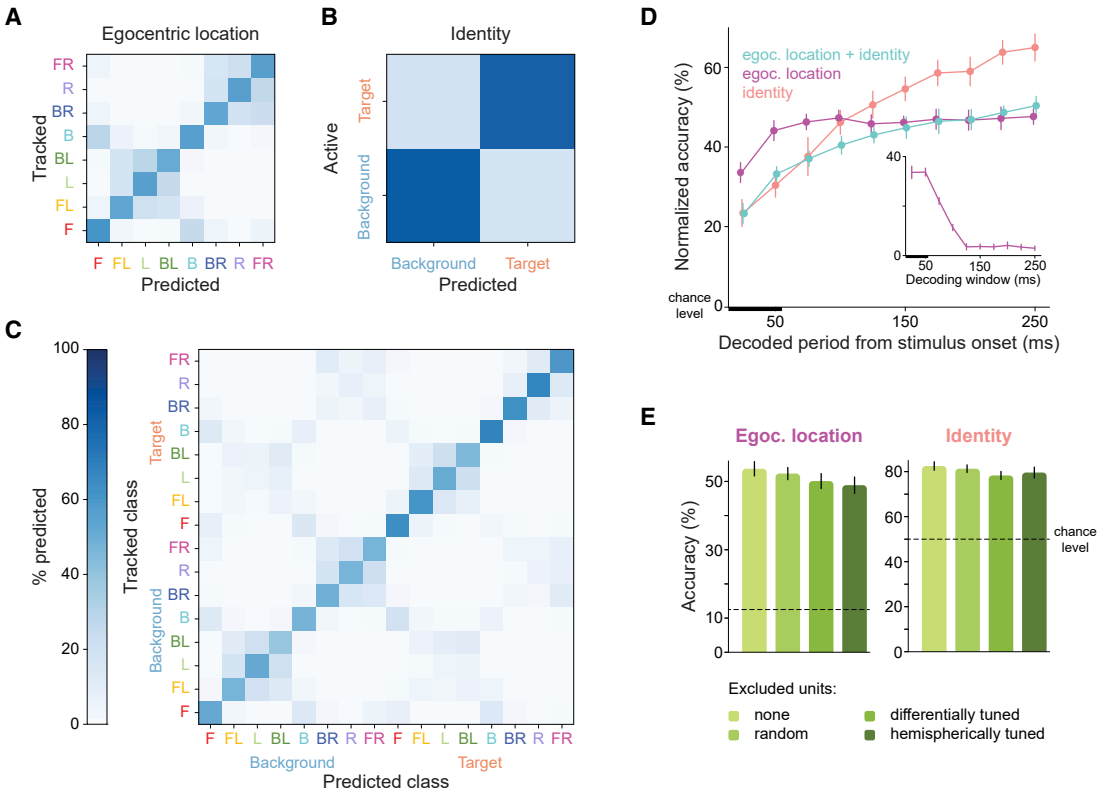

Figure 4. Artificial neural network decoding of population activity

(A–C) Confusion matrix for decoding (A) the egocentric loudspeaker location, (B) the identity of the active loudspeaker, and (C) the combination of the loudspeaker location and identity (classes). The angular bins used for decoding the egocentric loudspeaker location are the same as depicted in Figure 1C: FR, front right; F, front; FL, front left; L, left; BL, back left; B, back; BR, back right; R, right. For each tracked location/identity/class, the prediction probability was calculated for each predicted location/identity/class and is color-coded according to the color bar depicted in (C). For the construction of all the confusion matrices, 38,980 tracked/predicted classes were used across 20 sampling cycles (1,949 classes per cycle). See Results for decoding accuracy numbers.

(D) Normalized cumulative accuracy of the decoder as a function of the length of the time interval after sound onset used for decoding the egocentric location of the loudspeaker (as in A) and/or the identity of the loudspeaker (as in B). The solid black horizontal lines at the bottom of each plot represent the sound stimulation period. For each data point, the accuracy of 38,980 tracked/predicted classes was averaged across 20 sampling cycles (1,949 classes per cycle from both loudspeakers and all possible egocentric angles; STAR Methods). Inset: normalized accuracy for decoding the egocentric location of the active loudspeaker using a single 25 ms bin as a function of the time delay of the bin after sound onset. The normalized accuracy drops linearly from over 30% above chance level at 50 ms after sound onset to almost chance level after 125 ms after sound onset. All error bars correspond to the standard deviation of the mean normalized accuracy across 20 sampling cycles.

(E) Eliminating 18 units of different tuning types from the test dataset resulted in a minor drop in the decoding accuracy and remained highly significant above chance (accuracy relative to decoding without eliminating any units: differentially tuned, −3.6% ± 2.2% and −4.2% ± 1.8% for location and identity, respectively; randomly selected, −1.4% ± 1.8% and −1.2% ± 1.7%; hemispherically tuned, −4.8% ± 2.4% and −2.9% ± 2.4%). All error bars correspond to the standard deviation of the mean accuracy across 100 sampling cycles.

See also Figure S4.

interventions, as decoding accuracy for both reference frames remained far above chance level regardless of the eliminated neuronal tuning type (Figure S4). This robustness suggests that distributed interactions in population responses underlie an interlaced coding of egocentric and allocentric information.

Taken together, we demonstrate new cortical representations for active sensory object identification during unrestricted navigation. By introducing differential behavioral meaning to two loudspeakers at distinct locations, we found spatial representations in A1 neurons that were so far unreported. Maybe most intriguingly, we observed that the identity of the active sound source altered the spatial preference for physically identical stimuli. Specifically, neurons in the engaged A1 exhibit spatial tunings that combine ego- and allocentric (i.e., source identity) components to various degrees. This study is the first to report such an interlaced spatial code, possibly because the behavioral paradigm used is the first to provide simultaneous access to both reference frames (as in natural listening conditions). Likewise, the inclusion of natural self-motion renders pure egocentric coding insufficient (Figure 1E), which may have further favored revealing this novel coding regime. We also found a small subset of neurons whose response was modulated by the identity of the active sound source, but that were insensitive to egocentric locations. Since in our task source identities were distinguishable only by their respective allocentric positions, these neurons can be classified as purely allocentric. An earlier study[9] had also reported on a comparably small subset (6.5%) of A1 units with allocentric tuning in freely exploring but acoustically unengaged

ferrets. In stark contrast to our results, however, all units in their study were classified as either exclusively ego- or allocentrically tuned.

Sound-source identity also influenced the spatial tuning acuity, which was higher to the target compared to the background loudspeaker. Sharpening of spatial tuning in A1 has been previously reported for cats when engaged in a sound localization task compared to a periodicity task,[15] suggesting that our gerbils' localization behavior might have changed after entering the island.

Differential tuning to the two loudspeakers was especially apparent during late response periods (>~130 ms), which, interestingly, have been implicated to be crucial for the creation of auditory objects[24] and to encode perceived pitch and vowel identity in a discrimination task.[25,26] Although we did not directly probe the influence of other areas, the richness of our recorded late responses is indicative of cross-modal[27] and top-down modulation.[8,28] It thus seems likely that allocentric tuning aspects were conveyed to A1 by higher brain regions. Potentially relevant feedback connections originate in the retrosplenial cortex, which is strongly connected to the hippocampus, projects to A1,[29] and has been suggested to transform egocentric into allocentric information.[30] Allocentric representations were also found to be present in the cingulate cortex,[31] which also projects to A1.[29]

Irrespective of their spatial tuning, the neuronal response magnitude of A1 neurons differed based on the choice of the animals to remain in the island or to leave before receiving any reward. This finding is in line with earlier studies that found A1 neurons to encode task choice,[25,32–35] and could also be associated with predictive processing[36] and reward expectancy.[37] Ventral tegmental area is a possible source for reward signaling inputs via direct dopaminergic connections to infragranular layers of A1.[38] Such dopaminergic neurons have been shown to encode reward probability, and their response magnitude to predict changes in behavior,[39] which might be the mechanism underlying the decision of staying or leaving the island in our paradigm. Moreover, prefrontal cortex (PFC) has also been associated with encoding reward expectations,[40] as well as sound categorization.[41,42] Direct feedback in rodents from PFC to both A1[29] and the midbrain[43] renders this area also a likely candidate for top-down modulation in the computations described here.

Together, these novel neuronal computations appear well suited for identifying and tracking specific sound sources in the environment during self-motion. Moreover, while we did not present sounds from both sources simultaneously, our findings are consistent with prior observations of altered spatial sensitivity of A1 neurons in the presence of competing sources.[44] It would therefore be interesting to evaluate our results in the context of spatial stream segregation. More generally, we propose that experiments in freely moving and engaged animals are crucial to better understand long-standing phenomena of active sensing in everyday life such as scene analysis and the cocktail party problem.[45]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.06.025.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft DFG (PE2251/2-1 to M.P. and SFB 870, #118803580, project B02, to M.P. and B.G.), and by the International Max Planck Research School for Molecular Life Sciences (to D.A.). We thank N. Lesica and P. Alcami for valuable comments on earlier versions of the manuscript and H. Wohlfrom, O. Alexandrova, and K. Fischer for assistance with the histology. We are grateful to L. Wiegrebe for help on the behavioral paradigm.

AUTHOR CONTRIBUTIONS

M.P., D.A., and B.G. conceived the study. M.P. and D.A. designed the experiments and D.N.F. and B.G. contributed to paradigm refinement. D.A. performed the experiments and analyzed the data. D.A., M.P., and D.N.F. designed the data presentation and wrote the manuscript. All authors provided comments and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 3, 2021
Revised: May 10, 2021
Accepted: June 9, 2021
Published: June 29, 2021

REFERENCES

1. Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. Curr. Opin. Neurobiol. *20*, 172–176.
2. van der Heijden, K., Rauschecker, J.P., de Gelder, B., and Formisano, E. (2019). Cortical mechanisms of spatial hearing. Nat. Rev. Neurosci. *20*, 609–623.
3. Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. Physiol. Rev. *90*, 983–1012.

4. Mickey, B.J., and Middlebrooks, J.C. (2003). Representation of auditory space by cortical neurons in awake cats. J. Neurosci. *23*, 8649–8663.
5. Mrsic-Flogel, T.D., King, A.J., and Schnupp, J.W.H. (2005). Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. J. Neurophysiol. *93*, 3489–3503.
6. Stecker, G.C., Harrington, I.A., and Middlebrooks, J.C. (2005). Location coding by opponent neural populations in the auditory cortex. PLoS Biol. *3*, e78.
7. Woods, T.M., Lopez, S.E., Long, J.H., Rahman, J.E., and Recanzone, G.H. (2006). Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. J. Neurophysiol. *96*, 3323–3337.
8. Burgess, N. (2008). Spatial cognition and the brain. Ann. N Y Acad. Sci. *1124*, 77–97.
9. Town, S.M., Brimijoin, W.O., and Bizley, J.K. (2017). Egocentric and allocentric representations in auditory cortex. PLoS Biol. *15*, e2001878.
10. Benson, D.A., Hienz, R.D., and Goldstein, M.H., Jr. (1981). Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. Brain Res. *219*, 249–267.
11. Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat. Neurosci. *6*, 1216–1223.
12. David, S.V., Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. USA *109*, 2144–2149.
13. Lin, P.-A., Asinof, S.K., Edwards, N.J., and Isaacson, J.S. (2019). Arousal regulates frequency tuning in primary auditory cortex. Proc. Natl. Acad. Sci. USA *116*, 25304–25310.
14. von Trapp, G., Buran, B.N., Sen, K., Semple, M.N., and Sanes, D.H. (2016). A decline in response variability improves neural signal detection during auditory task performance. J. Neurosci. *36*, 11097–11106.
15. Lee, C.-C., and Middlebrooks, J.C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. Nat. Neurosci. *14*, 108–114.
16. Kato, H.K., Gillet, S.N., and Isaacson, J.S. (2015). Flexible sensory representations in auditory cortex driven by behavioral relevance. Neuron *88*, 1027–1039.
17. Ferreiro, D.N., Amaro, D., Schmidtke, D., Sobolev, A., Gundi, P., Belliveau, L., Sirota, A., Grothe, B., and Pecka, M. (2020). Sensory Island Task (SIT): a new behavioral paradigm to study sensory perception and neural processing in freely moving animals. Front. Behav. Neurosci. *14*, 576154.
18. Radtke-Schuller, S., Schuller, G., Angenstein, F., Grosser, O.S., Goldschmidt, J., and Budinger, E. (2016). Brain atlas of the Mongolian gerbil (Meriones unguiculatus) in CT/MRI-aided stereotaxic coordinates. Brain Struct. Funct. *221* (*Suppl 1*), 1–272.
19. Remington, E.D., and Wang, X. (2019). Neural representations of the full spatial field in auditory cortex of awake marmoset (*Callithrix jacchus*). Cereb. Cortex *29*, 1199–1216.
20. Zhou, M., Liang, F., Xiong, X.R., Li, L., Li, H., Xiao, Z., Tao, H.W., and Zhang, L.I. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. Nat. Neurosci. *17*, 841–850.
21. Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature *513*, 189–194.
22. Middlebrooks, J.C., Clock, A.E., Xu, L., and Green, D.M. (1994). A panoramic code for sound location by cortical neurons. Science *264*, 842–844.
23. Furukawa, S., and Middlebrooks, J.C. (2002). Cortical representation of auditory space: information-bearing features of spike patterns. J. Neurophysiol. *87*, 1749–1762.
24. Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? Biol. Cybern. *89*, 397–406.
25. Bizley, J.K., Walker, K.M., Nodal, F.R., King, A.J., and Schnupp, J.W. (2013). Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr. Biol. *23*, 620–625.
26. Town, S.M., Wood, K.C., and Bizley, J.K. (2018). Sound identity is represented robustly in auditory cortex during perceptual constancy. Nat. Commun. *9*, 4786.
27. Brosch, M., Selezneva, E., and Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J. Neurosci. *25*, 6797–6806.
28. Yin, P., Strait, D.L., Radtke-Schuller, S., Fritz, J.B., and Shamma, S.A. (2020). Dynamics and hierarchical encoding of non-compact acoustic categories in auditory and frontal cortex. Curr. Biol. *30*, 1649–1663.e5.
29. Budinger, E., and Scheich, H. (2009). Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. Hear. Res. *258*, 16–27.
30. Vann, S.D., Aggleton, J.P., and Maguire, E.A. (2009). What does the retrosplenial cortex do? Nat. Rev. Neurosci. *10*, 792–802.
31. Dean, H.L., and Platt, M.L. (2006). Allocentric spatial referencing of neuronal activity in macaque posterior cingulate cortex. J. Neurosci. *26*, 1117–1127.
32. Selezneva, E., Scheich, H., and Brosch, M. (2006). Dual time scales for categorical decision making in auditory cortex. Curr. Biol. *16*, 2428–2433.
33. Francis, N.A., Winkowski, D.E., Sheikhattar, A., Armengol, K., Babadi, B., and Kanold, P.O. (2018). Small networks encode decision-making in primary auditory cortex. Neuron *97*, 885–897.e6.
34. Guo, L., Weems, J.T., Walker, W.I., Levichev, A., and Jaramillo, S. (2019). Choice-selective neurons in the auditory cortex and in its striatal target encode reward expectation. J. Neurosci. *39*, 3687–3697.
35. Tsunada, J., Liu, A.S.K., Gold, J.I., and Cohen, Y.E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. Nat. Neurosci. *19*, 135–142.
36. Keller, G.B., and Mrsic-Flogel, T.D. (2018). Predictive processing: a canonical cortical computation. Neuron *100*, 424–435.
37. Poort, J., Khan, A.G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G.B., Mrsic-Flogel, T.D., and Hofer, S.B. (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. Neuron *86*, 1478–1490.
38. Brunk, M.G.K., Deane, K.E., Kisse, M., Deliano, M., Vieweg, S., Ohl, F.W., Lippert, M.T., and Happel, M.F.K. (2019). Optogenetic stimulation of the VTA modulates a frequency-specific gain of thalamocortical inputs in infragranular layers of the auditory cortex. Sci. Rep. *9*, 20385.
39. Babayan, B.M., Uchida, N., and Gershman, S.J. (2018). Belief state representation in the dopamine system. Nat. Commun. *9*, 1891.
40. Rouault, M., Drugowitsch, J., and Koechlin, E. (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. Nat. Commun. *10*, 301.
41. Jiang, X., Chevillet, M.A., Rauschecker, J.P., and Riesenhuber, M. (2018). Training humans to categorize monkey calls: auditory feature- and category-selective neural tuning changes. Neuron *98*, 405–416.e4.
42. Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. *13*, 1011–1019.
43. Olthof, B.M.J., Rees, A., and Gartside, S.E. (2019). Multiple nonauditory cortical regions innervate the auditory midbrain. J. Neurosci. *39*, 8916–8928.
44. Middlebrooks, J.C., and Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. J. Neurosci. *33*, 10986–11001.
45. Bizley, J.K., and Cohen, Y.E. (2013). The what, where and how of auditory-object perception. Nat. Rev. Neurosci. *14*, 693–707.
46. Bradski, G. (2000). The OpenCV Library. Dr. Dobbs J. Softw. Tools Prof. Program. *120*, 122–125.
47. Rolston, J., Gross, R., and Potter, S. (2009). Common median referencing for improved action potential detection with multielectrode arrays. In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1604–1607.

48. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Kenneth, D.H. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv. https://doi.org/10.1101/061481.

49. Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., et al. (2016). Spike sorting for large, dense electrode arrays. Nat. Neurosci. *19*, 634–641.

50. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11.

51. Hill, D.N., Mehta, S.B., and Kleinfeld, D. (2011). Quality metrics to accompany spike sorting of extracellular signals. J. Neurosci. *31*, 8699–8705.

52. Scargle, J.D., Norris, J.P., Jackson, B., and Chiang, J. (2013). Studies in astronomical time series analysis. VI. Bayesian block representations. Astrophys. J. *764*, 167.

53. Mazurek, M., Kager, M., and Van Hooser, S.D. (2014). Robust quantification of orientation selectivity and direction selectivity. Front. Neural Circuits *8*, 92.

54. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
UV curing dental cement	RelyX Unicem	Cat#640568
Medetomidin	Midas Pharma GmbH	CAS: 86347-14-0
Midazolam	Midas Pharma GmbH	CAS: 59467-70-8
Fentanyl	Midas Pharma GmbH	CAS: 437-38-7
Atipamezol	CP Pharma	CAS: 104054-27-5
Naloxon	Ratiopharm	CAS: 465-65-6
Flumazenil	Fresenius Kabi	CAS: 78755-81-4
Thilo-tears SE.	Alcon Pharma GmbH	N/A
Meloxicam, Metacam	Boehringer Ingelheim	CAS: 71125-38-7
Metamizole	Midas Pharma GmbH	CAS: 68-89-3
35% Orthophosphoric Acid, iBOND Etch 35 Gel	Kulzer	CAS: 7664-38-2
Green Fluorescent Nissl Stain NeuroTrace 500/525	Invitrogen	Cat#N21480
Non-Cyanide Gold Plating Solution	Neuralynx	N/A
Enrofloxacin, Baytril	Bayer	N/A
Paladur dental cement	Heraeus Kulzer	N/A
Deposited data		
Primary data	This paper	https://gin.g-node.org/dianamaro/Amaro_et_al_2021_CurrBiol
Experimental models: Organisms/strains		
Gerbils (*Meriones Unguiculatus*)	In-house breeding (Biocenter LMU-Munich)	N/A
Software and algorithms		
MATLAB	Mathworks	https://www.mathworks.com/
Python 3.5	N/A	https://www.python.org/
Spike sorting - Kilosort	N/A	https://github.com/cortex-lab/KiloSort
Bonsai	N/A	https://github.com/bonsai-rx/bonsai
Phy	N/A	https://github.com/cortex-lab/phy
Scikit-learn	N/A	https://scikit-learn.org/stable/
Multi Channel Experimenter	Multi Channel Systems	N/A
OpenCV	N/A	https://opencv.org/
Code used in the behavioral paradigm	N/A	https://gin.g-node.org/asobolev/runsit
Other		
Point Grey Flea Camera	FLIR Systems, OR	N/A
Arduino	N/A	https://www.arduino.cc/
Wireless headstage W2100-system	MultiChannel Systems	N/A
12.7 μm diameter tungsten 99.95% wire	California Fine Wire	N/A
Amplifier AVR 445	Harman/Kardon	N/A
Loudspeaker NSW1-205-8A 1" Extended Range	Aurasound	N/A

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Michael Pecka (pecka@bio.lmu.de)

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The generated dataset has been deposited to: https://gin.g-node.org/dianamaro/Amaro_et_al_2021_CurrBiol

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Gerbils

Experiments were conducted on 11 male Mongolian gerbils (*Meriones unguiculatus*) from the breeding colony of the Biocenter of the Ludwig-Maximilians University Munich. Only male individuals were used because of their larger weight, which facilitates carrying the implant. Sex-specific differences in localization behaviors or respective neural coding in A1 are not expected. Animals were housed in groups of 3 to 4 individuals at a temperature of 22.4°C and 66% humidity with 12 h light/dark cycles. Implanted individuals were housed individually. The experiments were conducted during the light phase of the cycle. All procedures were approved in accordance with the stipulations of the German animal welfare law (AZ 55.2-1-54-2532-74-2016).

METHOD DETAILS

Behavioral training and stimulus control

Behavioral data was collected from eleven gerbils. Animals were required to be at least 8 weeks old at the beginning of training and underwent a general habituation period in the SIT setup for 15 min per day for 5 days. Gerbils had unrestricted access to water in their home cages. Food was provided as pellets *ad libitum* until the training phase started, after which animals were only allowed food as rewards for successful trials (half of a sunflower seed or 20 mg, TestDiet LabTab AIN-76A). The weight of every gerbil was measured daily and additional food was provided if needed to ensure the weight not to drop more than 5% between consecutive training days and to maintain it within the desired range of 60–80 g.

Animals were trained using the SIT paradigm[17] for probing sensory perception in unrestrained and actively engaged animals. In SIT, a freely moving animal was trained to search for an unknown area in the arena (island), which prompted a change in the identity of the loudspeaker (from "background" to "target") from which a stimulation is played back, and to report the detection of the change by remaining in the island for a determined time interval ("SIT-time" = 6 s). The position of the island was random and changed every trial (pseudo-randomly chosen from a uniform distribution), making the loudspeaker change the only useful cue to find the island. After correctly reporting the detection of the target stimulus, the animal was rewarded with food that automatically dropped in the arena from an overhead food dispenser. Trials had a time limit of 60 s. If the animal did not correctly report the island within the time limit, a low-pass filtered noise was presented to the animal for 10 s, during which no new trial could be initiated.

Details of the setup and stimulation control were reported elsewhere[17] and are summarized here: The setup consisted of a circular arena (diameter = 92 cm) within a sound attenuated chamber (Figure 1A). Stimuli were computer generated and transmitted through an amplifier (AVR 445 Harman/Kardon, Germany) and small loudspeakers (Aurasound NSW1-205-8A 1" Extended Range) mounted externally of the arena (~5 cm distance to the perforated metal walls of the arena). Stimuli were 57 ms long harmonic complex sounds with a fundamental frequency of 147 ± 4 Hz and low-pass filtered below 1.5 kHz. Animals were trained to initiate a trial by remaining on an initiation platform (approx. 1 cm in height, 12 cm diameter) for 1 s. Trial initiation triggered the playback from the background loudspeaker (located either east or west of the initiation platform), and animal entrance into the island triggered the switch of the playback to the target loudspeaker (separated by 180° from the background loudspeaker). Stimuli were played at a repetition rate of 4 Hz and their amplitude was 70 dB SPL roved ± 5 dB. The animal's position was tracked via images captured every 250 ms with a Flea3 camera (FL3-U3-13Y3M-C, Point Grey Research). Stimulation parameters (loudspeaker identity) were updated online according to the animals' position within the arena. Custom-made software for animal tracking, stimuli generation and food reward delivery was developed in MATLAB.

No more than two training sessions were carried out per day, lasting up to 90 min. The training of the animals was performed by gradually reducing island size (starting at diameter = 42 cm, ~21% of the arena surface) and increasing SIT-time (starting with 2 s) over the course of the training sessions. The final parameters of the island size were diameter = 25 cm (~7% of the arena surface) and SIT-time = 6 s.

Video tracking

Custom-made online tracking scripts were developed in MATLAB. In the animals' tracking algorithm, an ellipse was fitted to the animal after background subtraction and the centroid of this ellipse was taken as the position of the gerbil. The recorded videos

were re-analyzed offline in Python (with OpenCV library[46]) where the ellipse, its centroid position and orientation were determined. The median orientation error was blindly assessed to be 6°, with 75% of the orientation errors being below 15°.

Behavioral data analysis

Data analyses were performed in MATLAB (Mathworks) and Python using custom scripts. To test the performance of the animals, we compared the percentage of successful trials in each session with surrogate runs based on random target island shuffling. That is, for each trial (offline, *a posteriori*), 1000 surrogate (non-real) islands, non-overlapping with the target one, were randomly set and the real trajectory of the animal was used to calculate in how many of these islands the trial would have been successful at each time point given the required SIT-time. The median chance performance and 95% confidence interval was calculated based on bootstrapping (random sampling with replacement from all the trials of the session). This method allows obtaining an estimate of the proportion of successful trials the animal would have gotten just by chance given their locomotion trajectory and dynamics.

Surgery

Electrophysiological data was collected from five trained gerbils. Animals were anesthetized with an intraperitoneal injection of a mixture of metedomidin (0.15 mg/kg), midazolam (7.5 mg/kg), and fentanyl (0.03 mg/kg). To maintain it at a constant level, the same mixture was subcutaneously re-injected every 90 min. After shaving and disinfecting the head, a local anesthetic (50 μl, 2% xylocaine) was injected under the scalp skin and below the skin near the ears. For protection and to prevent dehydration, the eyes were covered with an ophthalmic gel (Thilo-Tears SE, Alcon Pharma Gmbh). The animal was then transferred to the stereotactic apparatus, where its head was securely fixed via a bite and ear bars. Its internal temperature was monitored with a rectal thermometer and kept constant at 37°C throughout the experiment by a feedback controlled electric heating pad (Harvard Apparatus). After disinfection, a midline scalp incision was performed to expose the skull. Subsequently, the connective tissue on the skull was removed with a bone curette and the skull was treated with 35% phosphoric acid (iBOND etch gel, Kulzer), which was promptly washed away. Structural screws were placed on top of the left frontal and right parietal bones and the ground screw on the occipital bone, so that it gently touched the brain. After stereotactic alignment, a 3x3 mm craniotomy and durotomy were performed on top of the left auditory cortex, followed by a very slow lowering (2 μm/s) of a tetrode bundle to a maximum depth of 0.9 mm into the cortex, using a micromanipulator (Scientifica). The craniotomy was carefully filled with KY-jelly and immediately sealed with dental cement (Paladur, Kulzer), which also fixated the bottom of the microdrive and the outer cannula that protected the tetrodes. 1 mL of Ringer's solution was subcutaneously injected at the end of the surgery and the anesthesia was reversed via subcutaneous injection of the antagonist mixture composed of naloxone (0.5 mg/kg), flumazenil (0.4 mg/kg), and atipamezol (0.375 mg/kg). Analgesics (0.2 mg/kg, meloxicam) and antibiotics (7.5 mg/kg, enrofloxacin) were orally administered post surgically for five subsequent recovery days. During this time, the animals had food and water *ad libitum* and were not trained.

The implant used in this experiment was a tetrode bundle consisting of four tetrodes glued together, which, on their turn, consisted of four insulated tungsten wires (12.7 μm diameter each, tungsten 99.95%, California Fine Wire) twisted around each other. Each wire was connected to a custom-made printed circuit board with Omnetics connector (Axona), which was attached to a lightweight microdrive (0.25 mm/turn, Axona). The tetrodes were glued together and protected by an inner and outer cannula that could slide by each other. On the day prior to the surgery, the tip of all electrodes were cut with sharp scissors and gold plated (Non-Cyanide Gold Plating Solution, Neuralynx) to reach a desired impedance of 100-150 kOhm (at 1 kHz). The tetrode bundle was implanted vertically in the following coordinates from lambda: 6.2 mm lateral, 2.6 mm anterior.

Histology

After a 350 μl intraperitoneal injection of pentobarbital sodium, the animal was perfused with 4% paraformaldehyde (PFA) and the brain carefully removed and stored in PFA in the fridge. Afterward the brain was twice washed in PBS and the frontal part cut with a vibrotome in 70μm thick slices. The slices were stained with green fluorescent Nissl (NeuroTrace 500/525) and then compared to the gerbil brain atlas[18] for determining the location of the recording sites.

Electrophysiological recordings during task performance

Recordings were made via a wireless headstage (W2100-HS16, Multichannel Systems). The physiological signals were at first amplified between 1 Hz and 5 kHz and digitized (16-bit resolution) in the headstage, then wirelessly transmitted to the receiver (W2100-RE-AO, Multichannel Systems) at a sampling rate of 25 kHz and recorded by a PC via an interface board (MCS-IFB 3.0 Multi- boot, Multichannel Systems), and commercial software (Multi Channel Experimenter, Multichannel Systems). Simultaneous with the onset of sound presentation, a short signal was sent via sound card to the analog input of the interface board, allowing the synchronization of the physiological recordings with the sound presentation, and a digital signal was transmitted to the interface board indicating the beginning and end of the trial, which was later used to align the video information.

Spike sorting

Initially, the raw electrophysiological signals were high-pass filtered above 300Hz and a common median referencing[47] was performed for the whitening among all channels and removal of large artifacts. The signals were low-pass filtered below 5 kHz and fed to a spike sorting algorithm based on template matching (Kilosort[48]). Afterward, the automatically sorted spikes were manually inspected and the corresponding clusters refined with the graphical user interface phy.[49] Only units with an isolation distance larger

than 20, more than 200 spikes and less than 2% of the spikes within the 2 ms refractory period were considered single cells.[50,51] All units that passed the manual curation but not the isolation distance or refractory period requirements were considered multiunits.

Peristimulus time histograms and response period analysis

The peristimulus time histogram (PSTH) was calculated based on bootstrapping data. An algorithm sampled 500 times (with replacement) from the total number of recordings of a given sound. For each cycle, a histogram was calculated with binning width 8 ms and the median and 95% confidence intervals were determined for the total of bootstrapping cycles normalized by the total number of analyzed sounds. For the calculation of neuronal latencies without contamination by baseline activity, the Bayesian blocks method[52] was used, which identifies statistically significant variation in a time series to optimally segment the data. If at least one new edge is created, the unit is considered modulated and therefore auditory responsive. The latency of a given unit was defined as the first created edge. To determine whether the onset response corresponds to an increase in the firing rate, the maximum displacement of the median values of the firing rate in the 15 ms after latency was compared to the baseline (last 50 ms before latency). If it was lower than the mean of the lower bound of the baseline's 95% confidence interval, the onset firing rate was considered to be decreasing and if it was higher than the mean of the upper bound of the baseline's 95% confidence interval, the onset firing rate was considered to be increasing. The same Bayesian blocks method was employed to calculate whether a unit had a significant offset response within the interval [latency+50:latency+70] ms. If so, the offset period was defined as the time point of the new edge creation, otherwise it was defined as latency + sound duration (57 ms). The start of the late response period was defined as the start of the offset period + sound duration (57 ms).

Mean firing rates were calculated as the mean of the PSTH across all time bins.

For the comparison between the first and last 2 s inside the island and whenever the animal wrongly left it, only situations in which the animal stayed at least 2 s in the island were considered. Furthermore, if the last 2 s overlapped in time with the first 2 s (the animal stayed less than 4 s in the island), only the non-overlapping time period was used for the calculation of the last 2 s.

Spatial tuning analysis

For the spatial tuning analysis, the 360° angular space around the animal was divided in 8 bins. A minimum of 10 sound presentations per binned angle was a primary condition for the spatial tuning analysis, as well as a maximum difference of 10 between the bins with the most and the least number of sound presentations. A 1000 cycle bootstrapping method with replacement was implemented to calculate the spatial tuning, in which an angular histogram was calculated for the corresponding spikes and normalized by the number of sounds at each egocentric sound-source angle bin for the chosen loudspeaker. The vector strength and corresponding preferred egocentric sound-source angle were then calculated for each bootstrapping cycle. As some units revealed a more complex spatial tuning (with two peaks in opposite directions), the folded vector strength (in which responses to angles opposite to each other are summed[53]) was additionally calculated together with the corresponding orientation angle.

$$\rho = \frac{\sum_{j=1}^{8} \alpha_j \cdot e^{i\omega_j}}{\sum_{j=1}^{8} \alpha_j}$$

where $i = \sqrt{-1}, \alpha_j = \frac{\#\text{spikes in angular bin}_j}{\#\text{sound presentations in angular bin}_j}$ and $\omega_j = \frac{2\pi.\text{middle of angular bin}_j}{T}$

The vector strength is $VS = |\rho|$ with T = 360 and the corresponding preferred egocentric sound-source angle $ang_{dir}=arg(\rho)$. The folded vector strength is $VS_{folded} = |\rho|$ with T = 180 and the corresponding preferred egocentric sound-source angle angle $ang_{ori}=arg(\rho)/2$. From the bootstrapped data, the 95% confidence interval for each angle bin was calculated, as well as the 68% confidence interval of the vector strength and orientation vector. A unit was considered spatially tuned at a particular temporal period (onset, offset or late) if the vector strength (or folded vector strength, whichever is larger) was larger than 0.2 and the lower bound of 68% confidence interval of the vector strength was larger than 0.15.

Units were considered as untuned if the median vector strength was smaller than 0.15 in all three phases of the PSTH for both loudspeakers. The firing rate in response to the two loudspeakers was considered significantly different if there was no overlap in the 95% confidence interval of the calculation of the mean firing rate for each loudspeaker.

Decoder Analysis

For the decoder analysis, the decoder was trained on the recorded responses of 224 units (141 single neurons and 83 multi-units) from two animals in a total of 21 sessions. We chose to perform this analysis with the minimal number of animals and sessions to avoid overfitting. We found that data from 2 gerbils is sufficient to give robust results. Each sound presentation was categorized with respect to the active loudspeaker and the egocentric sound-source angle at the time of sound onset (in 8 angular bins). Therefore, each sound presentation corresponded to one of the 16 possible loudspeaker-angular bin combinations (classes). Spike counts, for each recorded unit, in each of the 250 ms response periods were determined in 10 bins, each with a 25 ms duration, and then normalized between 0 and 1 (using the mix-max scaler algorithm) to avoid that units that are inherently more active dominate the results. These unit responses were then pooled across sessions for each of the classes, creating 16 across-sessions class-specific population-response profile, which were then used to train the decoder. To prevent class imbalance, due to non-uniform occurrence of classes during the task, random under-sampling (without replacement) was implemented to feed the decoder with the same amount of data per class. The test dataset consisted of a minimum of 25% of the population responses per class.

The training data was fitted by a multi-layer perceptron classifier (MLPC[54]), a feedforward artificial neural network which uses backpropagation during training for optimization of the parameters in a supervised learning manner, and therefore was interpreted as a more biologically inspired decoder. In the MLPC, one hidden layer was implemented with the number of nodes as the mean between the number of features used and the total number of classes (16). The optimization algorithm used was the ''lbfgs'' from the family of quasi-Newton methods and the biologically-inspired rectified linear unit was the activation function for the hidden layer.

The whole process was repeated 20 times to estimate errors (100 times in the case of the unit-elimination analysis). The accuracy of identifying the active loudspeaker was determined per loudspeaker for each sampling cycle, later the accuracy per repetition was considered as the mean between the two loudspeakers and the total accuracy and standard deviation calculated across all 20 sampling cycles. The accuracy of predicting the right class and egocentric sound-source location was similarly calculated. The normalized accuracy was determined according to the formula: $\frac{accuracy(\%)-{}^{100}/_{N_{bins}}}{1-{}^{1}/_{N_{bins}}}$, where N_{bins} corresponds to the number of possible outcomes of the prediction (16, 8 and 2 for decoding the location-loudspeaker combination classes, only the egocentric sound-source location or only the identity of the active loudspeaker, respectively). Consequently, a chance level model performance corresponds to 0% normalized accuracy and a 100% performance to 100% normalized accuracy. For the construction of the confusion matrices, all the predictions from all the sampling cycles were simultaneously used.

To evaluate the influence of units with different spatial tuning, units were eliminated in the test data by setting all the bins corresponding to the chosen units to zero. The accuracy of the decoder (trained on the training data from all the units) was then compared to when the test data was complete. In the elimination of the random units, different units were chosen to be eliminated in each sampling turn. Elimination of ''differentially tuned'' units included all units in the training set which at some temporal response period had a difference in preferred egocentric sound-source angle between loudspeakers larger than 80°. The eliminated canonical units were randomly selected from the neurons which showed clear ipsi or contralateral tuning during the onset period.

QUANTIFICATION AND STATISTICAL ANALYSIS

All error bars correspond to the standard deviation unless stated otherwise. All shaded areas correspond to the 95% confidence interval as calculated via bootstrapping unless stated otherwise. For comparisons of central tendencies on the group level, we used two-tailed non-parametric tests: Wilcoxon signed-rank tests for paired samples and Mann-Whitney U-test for independent samples. The data presented was non-uniformly distributed, and thus we used non-parametric testing and visualization methods. All hypotheses were tested at an alpha level of 0.05.

Data collection and analysis were not performed blind to the conditions of the experiments. Our experimental design provided a within-animal control, and because comparisons were not required between different groups, blinding was not necessary. All analyses were based on automated scripts applied across animals and thus were not subject to any experimenter bias.