

RESEARCH ARTICLE

Computational Neuroscience

A robust and compact population code for competing sounds in auditory cortex

 Jian Carlo Nocon,^{1,2,3,4}  Jake Witter,⁵ Howard Gritton,^{6,7} Xue Han,^{1,2,3,4}  Conor Houghton,⁵ and
 Kamal Sen^{1,2,3,4}

¹Neurophotronics Center, Boston University, Boston, Massachusetts, United States; ²Center for Systems Neuroscience, Boston University, Boston, Massachusetts, United States; ³Hearing Research Center, Boston University, Boston, Massachusetts, United States; ⁴Department of Biomedical Engineering, Boston University, Boston, Massachusetts, United States; ⁵Department of Computer Science, University of Bristol, Bristol, United Kingdom; ⁶Department of Comparative Biosciences, University of Illinois, Urbana, Illinois, United States; and ⁷Department of Bioengineering, University of Illinois, Urbana, Illinois, United States

Abstract

Cortical circuits encoding sensory information consist of populations of neurons, yet how information aggregates via pooling individual cells remains poorly understood. Such pooling may be particularly important in noisy settings where single-neuron encoding is degraded. One example is the cocktail party problem, with competing sounds from multiple spatial locations. How populations of neurons in auditory cortex code competing sounds have not been previously investigated. Here, we apply a novel information-theoretic approach to estimate information in populations of neurons in mouse auditory cortex about competing sounds from multiple spatial locations, including both summed population (SP) and labeled line (LL) codes. We find that a small subset of neurons is sufficient to nearly maximize mutual information over different spatial configurations, with the labeled line code outperforming the summed population code and approaching information levels attained in the absence of competing stimuli. Finally, information in the labeled line code increases with spatial separation between target and masker, in correspondence with behavioral results on spatial release from masking in humans and animals. Taken together, our results reveal that a compact population of neurons in auditory cortex provides a robust code for competing sounds from different spatial locations.

NEW & NOTEWORTHY Little is known about how populations of neurons within cortical circuits encode sensory stimuli in the presence of competing stimuli at other spatial locations. Here, we investigate this problem in auditory cortex using a recently proposed information-theoretic approach. We find a small subset of neurons nearly maximizes information about target sounds in the presence of competing maskers, approaching information levels for isolated stimuli, and provides a noise-robust code for sounds in a complex auditory scene.

auditory cortex; complex scene analysis; information theory; neural coding; population coding

INTRODUCTION

A central, surprising finding of systems neuroscience is that the discrimination performance of single cortical neurons can match behavior (1, 2). However, some outstanding questions are whether single neurons can withstand highly noisy settings, and whether population coding, in which the activity of multiple neurons are aggregated to represent stimuli, can improve discrimination performance in such settings. An important example of a noisy setting is the cocktail party problem, where competing sounds originate from different spatial locations (3).

Such settings are highly challenging for a variety of populations with impairments, for example, attention-deficit/hyperactivity disorder (ADHD), autism and hearing impairment, assistive devices, and for speech recognition technology, for example, SIRI and Alexa. Understanding how information is coded in such noisy settings by single neurons and how it aggregates with population coding may illuminate better strategies for treatments in impaired populations, as well as improvements in assistive devices and speech recognition in noise.

Previous studies in songbirds quantified the discrimination performance of single neurons and found degraded

discrimination of target sounds in the presence of a competing masker at the same location (4), with high performance levels when the target and masker are spatially separated (5). In this case, the best single neuron code may suffice to support behavior. However, a more recent study in the mouse auditory cortex (ACx) with a similar experimental design found that discrimination performance of single neurons in the presence of competing sounds is significantly degraded in the presence of competing sounds (6). In this case, population coding may be necessary to support behavior.

Recent studies in ACx have investigated population coding of natural sounds (7) and dynamic amplitude-modulated sounds (8). However, population coding of competing sounds has yet to be investigated, motivating the following questions: can population coding improve the representation of competing sounds? How do different coding schemes compare? What is the size and composition of the best population under such schemes? Here, we investigate these questions in mouse ACx by applying a novel information-theoretic approach. We examine these questions using two population coding schemes: the summed population (SP) code, where the identity of each neuron does not impact coding, and the labeled line (LL) code, where neuron identity is preserved and can impact coding (9).

Many information-theoretic approaches require binning spike trains at a certain temporal resolution (7), whereas the relevant temporal resolution in cortex is unknown. Other approaches have used spike distance metrics that are free of a choice of temporal resolution (10, 11). However, spike distance-based approaches have typically been used in conjunction with a specific classifier (2, 4–6). It is unclear how to choose a classifier that best corresponds to cortical computations, or how the choice of classifier influences estimates of discrimination performance. Here, we combine the strengths of a time scale-free distance measure (10, 11), and a recently proposed classifier-free information-theoretic approach (12, 13) to probe population coding of competing sounds in mouse ACx. We find that population coding achieves near-maximal information levels with a surprisingly compact population of neurons in both the SP and LL schemes. Furthermore, the LL scheme outperforms the SP scheme, using a population of neurons with diverse spatial responses and cell types, greatly improving information available from single neurons, and approaching information levels of “clean” stimuli without competing noise. Finally, information available in population in the LL scheme increases with spatial separation between competing sounds, matching spatial release from masking observed at the behavioral level in animals and humans. Our results reveal a robust and compact, population code for competing sounds in ACx.

MATERIALS AND METHODS

Subjects

All procedures involving animals were approved by the Boston University Institutional Animal Care and Use Committee and the University of Illinois at Urbana-Champaign Institutional Animal Care and Use Committee (IACUC). A total of 12 mice were used in this

study. Original breeding pairs of parvalbumin-Cre (PV-Cre: B6;129P2-Pvalbtm1(cre)Arbr/J) and Ai40 mice (Arch: B6.Cg-Gt(ROSA)26Sor^{tm40.1(CAG-aop3/EGFP)Hze/J}) mice were obtained from Jackson Laboratory (Maine), and all breeding was done in-house. Subjects consisted of both male and female transgenic PV-Arch ($n = 7$ mice) offspring and PV-Cre ($n = 5$ mice) only offspring 8- to 12-wk old on the day of recording.

Surgery

Under isoflurane anesthesia, stereotaxic surgery was performed on mice to install a head plate, electrode, and optical fiber (14, 15). The custom head-plate was mounted anterior to the bregma to allow caudal access to ACx and anchored to the skull with three stainless steel screws and dental cement. A fourth screw was connected to a metal pin and placed in the skull above the contralateral cerebellum to serve as the reference. A craniotomy was made above the right auditory cortex [anteroposterior (AP) -2.3 to -3.6 , mediolateral (ML) $+4.0$ to $+4.5$, dorsoventral (DV)]. Using a stereotaxic arm, a 32-contact linear probe (Neuronexus, Ann Arbor, MI; Model: a 4×8 -5 mm-100-400-177-CM32) with 100- μm spacing between electrode contacts and 400- μm spacing between shanks was positioned into ACx, perpendicular to the cortical surface. Because of the curvature of the ACx surface, not all four shanks could be placed at precisely the same depth during each experiment. Probes were advanced until all electrode contacts were within the cortical tissue and shanks were positioned along the rostrocaudal axis of ACx (Fig. 1, A and B). An optical fiber, 200 μm in diameter, was placed medially to the four shanks and positioned between the two innermost shanks terminating at the cortical surface (Fig. 1A and Supplemental Fig. S1). After implantation, mice were allowed to recover for 4–7 days before undergoing habituation to being head-fixed as described in the following section.

Habituation

Following surgery and complete recovery, mice were first handled for several days before being head-fixed to the recording apparatus. Mice were gradually exposed to longer restraint periods at the same time of day as subsequent recording sessions and received at least six habituation sessions before the first recording day. Under head-fixed conditions, mice were loosely covered with a piece of laboratory tissue taped down on either side (Kimwipes: Kimberly-Clark, Irving, TX) to encourage reduced movement. At the end of habituation, mice underwent recording sessions in the presence of spatially distributed auditory stimuli.

Auditory Stimuli

All auditory stimuli were generated in Matlab and consisted of either target, masker, or combination of the two stimuli played from four TDT ES-1 electrostatic speakers. Target stimuli consisted of white noise modulated in time by human speech envelopes taken from the Harvard IEEE speech corpus (16), which has been used in previous psychological studies of the cocktail party effect (17). Masker stimuli consisted of 10 independent tokens of unmodulated white noise. All stimuli were 3-s long, with a 1 ms cosine ramp

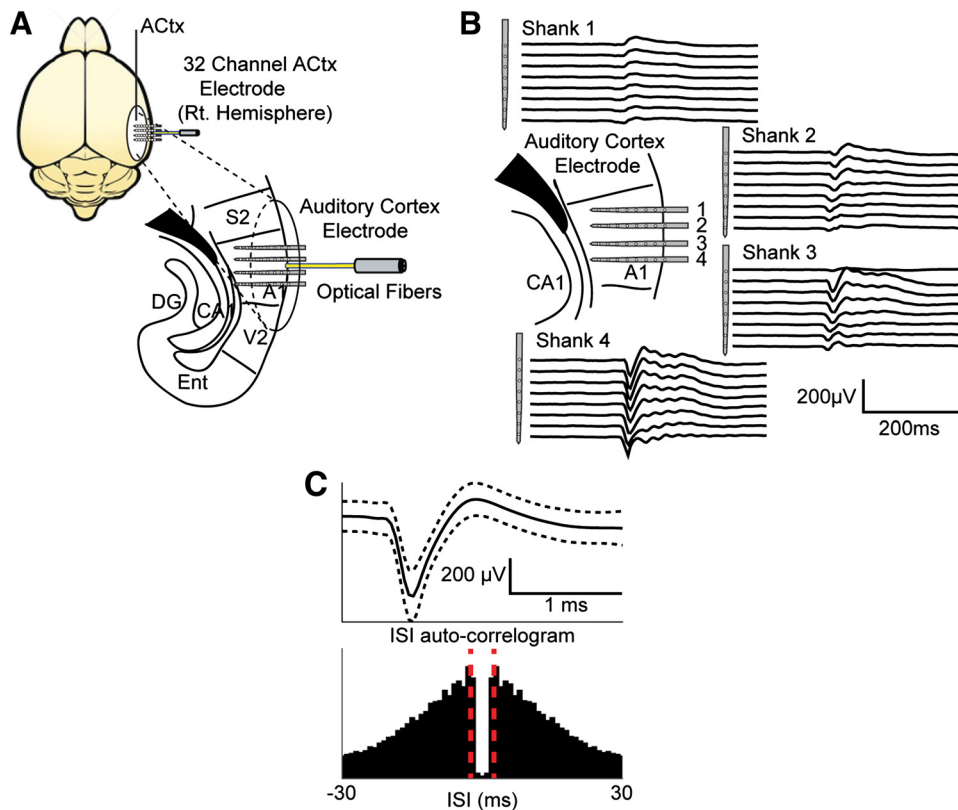


Figure 1. Experimental methods. All subplots replicated from Nocon et al. (6), licensed under a Creative Commons license. **A:** recording electrode location and optical fiber placement. Subjects were implanted with a 4-shank, 32-channel electrode array and optogenetic fiber in the right hemisphere of ACx. Each shank contained eight sites per shank with 100 μm spacing between electrode contacts. Mouse brain illustration is from Pixta (<https://www.pixtastock.com/illustration/67155575>). A1, primary auditory cortex; ACx and ACtx, auditory cortex; CA1, CA1 field; DG, dentate gyrus; Ent, entorhinal area; S2, secondary somatosensory cortex; V2, secondary visual cortex. **B:** representative local field potential (LFP) activity from one subject. LFPs were used to estimate current source density and the granular layer (L4) within each shank. **C:** example mean single-unit waveform and interspike interval (ISI) autocorrelogram. Black dashed lines in the mean waveform represent ± 1 SD, while scale bars measure voltage and time. Dashed red lines in the correlogram represent ISIs of ± 2 ms.

at onset and offset. Stimuli were loaded onto a custom RPDsEx circuit on an RZ6 Multi I/O processor, which was connected to two PM2R multiplexers that controlled the spatial location of stimuli during playback.

Recording Sessions and Data Acquisition

All recordings were made with a Tucker Davis Technologies (TDT; Alachua, FL) RZ2 recording system in an electrically shielded sound attenuation chamber. Broadband neural signals at 24,414.0625 Hz were recorded for each of the 32 channels. Local field potentials (LFPs) were band-pass filtered between 1 and 300 Hz, notch-filtered at 60 Hz, and digitized at 3,051.8 Hz and used for current source density analysis.

Recording sessions consisted of both nonoptogenetic and optogenetic trials in random order. The intertrial interval was 5 s, with 3 s of stimulus playback followed by 2 s of silence. Mice were exposed to target-alone (clean) trials and target-masker (masked) combinations. Ten trials were given per target identity for all possible combinations of target location, masker location (including clean trials), and optogenetic suppression of PV neurons. Thus, animals received a total of 800 trials per ~ 60 -min recording session, with each session having a set laser power. For the present study, we focus on the responses recorded during nonoptogenetic trials.

During recordings, the stimuli were presented 18 cm from the mouse's head using four TDT ES-1 speakers driven by two TDT ED-1 speaker drivers. The four speakers were arranged around the mouse at four locations on the azimuthal plane: directly in front (0°), two contralateral (45° and 90°), and one ipsilateral (-90°) to the right auditory

cortex recording area. Before recording sessions, stimulus intensity was calibrated using a conditioning amplifier and microphone (Brüel and Kjær, Nærum, Denmark; Amplifier Model: 2690, and Microphone Model: 4939-A-011). For seven of the nine Arch mice and the five Arch-nonexpressing animals, all stimuli were set to 75 dB intensity at the mouse's head. For the remaining two Arch mice, stimulus intensity was set to 70 dB.

Spike Extraction and Clustering

Kilosort2.0 (<https://zenodo.org/record/4147288>) was used to automatically detect single units (SUs) (18), with the broadband signal passed through a third-order Butterworth filter with band-pass frequencies 300 and 5,000 Hz. Kilosort results were then loaded onto Phy2 (<https://github.com/cortex-lab/phy>) to manually determine if spike clusters exhibited neural activity or noise (19). Clusters with either artifact-like waveforms from laser activity or similar responses across all channels were rejected, and spikes with artifact-like waveforms were removed from clusters whenever possible. Clusters were merged if the cross-correlograms were similar to the component clusters' autocorrelograms and showed overlap in principal component feature space at the same channel. The spikes toolbox (<https://github.com/cortex-lab/spikes>) was used to import the cluster information from Phy2 to Matlab and extract spike waveforms from the high-passed signal (18). Clusters were assigned to recording channels based on which site yielded the largest average spike amplitude. To remove any remaining artifacts from laser onset and offset, all spikes with waveforms above an absolute threshold of 1,500 μV or a positive value above 750 μV

were discarded, and clusters that still showed a high amount of remaining artifact-like spikes after removal were excluded from further analysis. To determine which of the remaining clusters were SUs, we used the sortingQuality toolbox (<https://github.com/cortex-lab/sortingQuality>) to calculate isolation distances and L-ratios (20). SUs must 1) have less than 5% of interspike intervals below 2 ms (Fig. 1C), 2) an isolation distance above 15, and 3) an L-ratio below 0.25. For clusters where isolation distance and L-ratio were not defined, only the first threshold was used. These thresholds are consistent with past studies on single-unit activity (21–23), and clusters that did not meet any of these criteria were deemed multi-units (MUs). Finally, SUs were classified as narrow-spiking if the trough-peak interval of their mean waveform was below 0.5 ms, which is consistent with past studies on inhibitory units within mouse auditory cortex (24).

Current Source Density Estimation and Layer Analysis

Current source density (CSD) analysis estimates the second spatial derivative of LFP signals to determine the relative current across the cortical laminar depth. Similar to previous studies (15), LFPs were first low-passed filtered at 150 Hz before being downsampled by a factor of 8 to 381 Hz. For each channel, LFPs were averaged across all control masked trials before CSD estimation, as the rise time was more similar between target stimuli than in clean trials. Channels that did not show an evoked response were interpolated using neighboring sites on the same shank. LFPs (Φ) were then spatially smoothed across the eight channels in each shank:

$$\phi(z) = \frac{\phi(z + \Delta z) + 2\phi(z) + \phi(z - \Delta z)}{4} \quad (1)$$

where z is the depth perpendicular to the cortical surface and Δz is the electrode spacing. CSD was then estimated as:

$$\text{CSD}(z) = -\frac{\phi(z + \Delta z) - 2\phi(z) + \phi(z - \Delta z)}{\Delta z^2} \quad (2)$$

For each of the middle six channels, CSD sink onset was calculated as the time when the CSD goes below -3 times the standard deviation of prestimulus activity. If more than one channel was found to have the earliest sink onset, the channel whose neighbors yielded the smallest difference in onset time was deemed the granular layer, or L4. The depths of each layer were estimated based on previous anatomical studies (25): L1 consisted of channels at least 500 μm above the input layer, L2/3 consisted of channels 200–400 μm above the channel with the earliest sink onset; L4 consisted of the input channel and the channel 100 μm above it; L5 consisted of channels 100–300 μm below the input layer, and L6 consisted of all channels at 400 μm below the input layer.

Neural Discriminability Performance Using SPIKE-Distance

Neural discrimination performance refers to the ability to determine stimulus identity based on neural responses, thus measuring a neuron's ability to encode stimulus features. Here, performance was calculated using a template-matching approach similar to our previous studies (5). Spike trains were compared to template responses from both target stimuli, and each train was classified as being elicited by the

target whose template yielded the smaller spike distance. For each target-masker configuration, 100 iterations of template matching were done. In each iteration, one of the 10 spike trains for each target was chosen as a template, and all remaining trials were matched to each template to determine target identity. All possible pairs of templates were used across the 100 iterations to calculate an average value of neural discriminability. SPIKE-distance (10) calculates the dissimilarity between two spike trains based on differences in spike timing and instantaneous firing rate without additional parameters. For one spike train in a pair, the instantaneous spike timing difference at time t is:

$$S_1(t) = \frac{\Delta t_P^{(1)}(t)x_F^{(1)} + \Delta t_F^{(1)}(t)x_P^{(1)}}{x_{\text{ISI}}^{(1)}(t)}, t_P^{(1)} \leq t \leq t_F^{(1)} \quad (3)$$

where Δt_P represents the distance between the preceding spike from train 1 [$t_P^{(1)}$] and the nearest spike from train 2, Δt_F represents the distance between the following spike from train 1 [$t_F^{(1)}$] and the nearest spike from train 2, x_F is the absolute difference between t and $t_F^{(1)}$, x_P is the absolute difference between t and $t_P^{(1)}$, and $x_{\text{ISI}}^{(1)}$ is the interspike interval (ISI) between t_F and t_P in spike train 1. All these quantities are functions of t , as they measure the distances between spikes and the current time t or between spikes whose identity will change with t , as illustrated in Supplemental Fig. S2, A and B. To calculate $S_2(t)$, the spike timing difference from the view of the other train, all spike times and ISIs are replaced with the relevant values in train 2. The pairwise instantaneous difference between the two trains is calculated as:

$$S''(t) = \frac{S_1(t) + S_2(t)}{2\langle x_{\text{ISI}}^{(1)}(t), x_{\text{ISI}}^{(2)}(t) \rangle} \quad (4)$$

Finally, $S_1(t)$ and $S_2(t)$ are locally weighted by their instantaneous interspike intervals to account for differences in firing rate:

$$S(t) = \frac{S_1(t)x_{\text{ISI}}^{(2)}(t) + S_2(t)x_{\text{ISI}}^{(1)}(t)}{2\langle x_{\text{ISI}}^{(1)}(t), x_{\text{ISI}}^{(2)}(t) \rangle^2} \quad (5)$$

For a train of length T , the distance is the integral of the dissimilarity profile across the entire response interval, with a minimum value of 0 for identical spike trains:

$$D_S = \frac{1}{T} \int_0^T S(t) dt \quad (6)$$

cSPIKE (<https://www.thomaskreuz.org/source-codes/cspike>), a toolbox used to calculate SPIKE-distance, was used to calculate all spike train distances between all possible spike train pairs for all target-masker configurations (10).

Identifying “Hotspots” of High Neural Discriminability

To restrict the number of single units used for population coding analysis, we used only the single units that showed at least one “hotspot” or target-masker configuration showing high neural discriminability during either clean or masked presentations. “Hotspots” were determined using three criteria: 1) performance must be above 70% during the control

condition; 2) the control performance distribution must be significantly different from chance ($P < 0.05$), calculated using a null distribution obtained by classifying spike trains within each target, which results in performances around chance value 50%; and 3) the effect size between the two distributions (control vs. null), calculated using Cohen's d , must be greater than 1:

$$d = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2}}} \quad (7)$$

where values with subscript 0 represent the mean, standard deviation, and number of template-matching iterations for the null performance distribution. In addition, configurations where at least three trials for one target showed zero spiking were excluded from analysis, to avoid inaccurate estimates of performance.

Of the 137 identified SUs, 45 exhibited at least one hotspot. To determine whether the remaining SUs were silent during stimulus coding, we measured the auditory responsiveness of all units. We compared the average and maximum firing rate during stimulus playback to spontaneous activity using unpaired Wilcoxon rank-sum tests. These two comparisons were done for each of the 20 target-masker configurations and two target identities, resulting in 80 total statistical tests per unit. An SU was deemed auditory-responsive if any of these tests were found to be significant after a Bonferroni correction ($P < 0.05/80$). Ninety-seven of the 137 SUs were found to be auditory responsive, including all 45 SUs with at least one hotspot. Thus, 52 of the remaining 97 SUs without a hotspot were found to be auditory-responsive. These findings are consistent with other studies that show a sparse-evoked population response within auditory cortex (26, 27), with single units exhibiting sustained responses for preferred stimuli and transient responses for nonpreferred stimuli (28). For the remaining analysis, we restricted our population of SUs to the 45 that exhibited at least one hotspot.

Rate-Normalized Root-Mean-Square Difference and Trial Similarity

To characterize the distribution of responses from our single units, we calculated two other measures: the similarity of responses within a given target and the dissimilarity of responses across targets. To quantify intertrial reliability of responses to target stimuli, we adopted the measure of trial similarity from previous studies (29). Specifically, we randomly divided the 10 trials in each configuration into two equal groups, binned spike times with a time resolution of 25 ms, and calculated the Pearson's correlation coefficient between the two resulting peristimulus time histograms (PSTHs). This process was repeated 100 times to obtain a mean correlation coefficient or trial similarity.

We also calculated the rate-normalized root-mean-square (RMS) difference between target responses to quantify the dissimilarity in the temporal pattern of responses between the two targets. We first binned each target response using the same time resolution as trial similarity (25 ms) and normalized each peristimulus time histogram (PSTH) such that the sum of all bins over time was 1. The RMS difference between the two rate-normalized PSTHs was then calculated. This

measure quantifies the dissimilarity in the temporal pattern of responses between both targets while accounting for differences in mean evoked firing rate.

Calculating Mutual Information Using Spike Train Distances

To calculate the mutual information of target identity from spike trains for each target-masker location configuration, we used an estimator on SPIKE-distances. From these distances, mutual information of stimulus identity was calculated using a Kozachenko-Leonenko estimator (12, 13, 30, 31), which estimates mutual information on a metric space. It is derived without any reference to a coordinate structure, something the space of spike trains lacks. Essentially, it estimates mutual information by approximating probability densities using probabilities:

$$p(x) \approx \int_B p(y) dy = \text{Prob}(x \in B) \quad (8)$$

for a region B . $\text{Prob}(x \in B)$ is then estimated by counting the number of data points in B . In this approach, the distance metric is used to define the small region and the estimator itself is a sort of nearest-neighbor formula, requiring us to look at which data points are near to each other. Calculating mutual information typically requires a lot of data, but this formula appears to work well even when there are not many data points and can be effectively debiased (32). Supplemental Fig. S2C illustrates the calculation of this measure of mutual information. To determine whether the layer location of single units influenced target discriminability, we ran a one-way ANOVA on the single-unit mutual information (MI) values across all target-masker configurations, including those from clean trials, with layer as the between-group factor. Post hoc multiple comparisons were done if the ANOVA yielded statistical significance ($P < 0.05$).

Population Searches Using Summed Population and Labeled Line Hypotheses

Given a full population of N encoding cells $X = [x_1, x_2, \dots, x_N]$, we wanted to determine the subset within X that best encodes stimulus information, which we define as K_{opt} with size $n < N$, where n is the number of single units in a subpopulation. For each target-masker configuration, searches for K_{opt} were carried out using either the summed population (SP) or labeled line (LL) approach. The SP code hypothesizes that mutual information is optimized by pooling various neurons to create a single, population-wide response. Specifically, the single response is a union of all spike times from each individual train, and coincident spike times from multiple units are only counted once (10). Meanwhile, the LL code hypothesizes that stimulus features are best decoded on a neuron-by-neuron basis. In this approach, responses from different units were concatenated in time to create a response whose length is the product of the trial length and the number of units. Unlike the SP code, the LL code keeps responses from each single unit intact. To account for the deletion of spike times in the SP code, all coincident spike times in the LL code were deleted such that only one neuron would spike at a given time. In both approaches, we treated all units as pseudo-simultaneous.

For all population searches, a bottom-up forward selection algorithm from a study by Satuvuori et al. (11) was used. The algorithm first starts with the single neuron that yields the best mutual information and then builds up the population by adding all remaining units at each step, based on the resulting mutual information (Fig. 3C). To account for unnecessary additions to the subpopulation due to plateaus in mutual information values, we define K_{opt} as the value of n that reached 90% of the maximum MI across all subset sizes. To determine how target-masker location configurations affect discriminability, the “upper envelope” was defined as the maximum MI at each target-masker configuration. To determine whether the MI value from K_{opt} was statistically different from null values, we shuffled the labels of each target and calculated the resulting MI for 1,000 permutations to create a 95% confidence interval. Statistical significance was yielded if the MI from the unshuffled trial labels was outside of the confidence interval. Finally, to measure the variance in MI for each neuron subpopulation, we ran a leave-one-out approach. Briefly, in each iteration, one trial from each of the two targets was removed before calculating SPIKE distances and estimating MI. All possible pairs of removed trials were used, resulting in a distribution of 100 MI values per configuration.

RESULTS

Investigating Cortical Coding in Mouse ACx Using a Cocktail Party-Like Paradigm

We recorded single units (SUs) and multi-units (MUs) using a multielectrode array with four shanks and 32 channels throughout different layers in ACx of unanesthetized PV-Arch transgenic mice (Fig. 1 and Supplemental Fig. S1). We used a semiautomated detection and sorting algorithm to identify 192 units from $n = 12$ animals (18, 19). Of these 192 units, 137 were identified as SUs (e.g., Fig. 1C). To better understand cortical coding of complex scenes, we adopted a cocktail party-like experimental paradigm (5) while recording from neurons in ACx. Specifically, we recorded responses to spatially distributed sound mixtures to determine how competing sound sources influence cortical coding of stimuli. The recording configuration consisted of four speakers arranged around the mouse at four locations on the azimuthal plane: directly in front (0°), two contralateral (45° and 90°), and one ipsilateral (-90°) to the right auditory cortex recording area. Target stimuli consisted of white noise modulated by human speech envelopes extracted from a speech corpus (16). We used two target waveforms (*target 1* and *target 2*) and a competing masker consisting of unmodulated white noise. Mice were exposed to either target-alone trials (Clean) or target-masker combinations (Masked) (Fig. 2, A–C).

Enhanced Neural Discriminability is Dependent on Spatial Configuration of Competing Auditory Stimuli

We assessed cortical coding using neural discriminability, which refers to the ability to determine stimulus identity based on neural responses and thus a neuron’s ability to encode stimulus features. Neural discriminability between the two targets (% correct) was computed for trials without

the masker (Clean) and with the masker (Masked) for all possible combinations of target and masker locations using a spike train distance-based classifier to determine how well target identity can be decoded based on dissimilarities between responses in spike timing and instantaneous rate (5). Values near 100% and 50%, respectively, represent perfect discriminability and chance discriminability. Configurations of target and masker locations with high neural discriminability were deemed as “hotspots” (6) using three criteria: 1) performance must be $\geq 70\%$; 2) statistically significant ($P < 0.05$) from a null distribution in which responses from a given target were template-matched with themselves, yielding chance values, and 3) effect size d must be greater than 1, relative to the null distribution. Figure 2A illustrates spike trains from an example SU that shows high discriminability under a target-only configuration and a specific spatial configuration in the presence of a competing noise masker (Fig. 2B). In another example configuration of target and masker location (Fig. 2C), discriminability is greatly reduced, indicating that the response of this neuron is spatial configuration sensitive.

Generally, discriminability between targets depends on the similarity of responses to the same stimulus as well as the dissimilarity in responses between different stimuli. To characterize both components in our population of high-performance SUs, we calculated the trial similarity within targets (Fig. 2D) and the rate-normalized RMS difference between targets (Fig. 2E). The distributions for each metric during clean and masked trials show that when a competing masker is introduced, both within-target reliability and between-target differences in single units decrease relative to clean presentations of target stimuli. From these results, a question arises: how can populations of single units aggregate their activity to restore neural discriminability in the presence of competing sound sources? Previous studies have demonstrated that neurons with the highest performance are most strongly correlated with behavior and strongly constrain population performance (1, 2, 8, 33, 34). For the following population coding results, we restricted our analysis to the 45 single units that showed at least one “hotspot.” Supplemental Table S1 shows the amount of SUs with at least one hotspot per layer and per animal.

Description of Experimental Data and Population Searches

Using the 45 single units that exhibited at least one “hotspot” of high neural discriminability in response to spatially distributed auditory stimuli, we determined the neural subpopulation that gives the highest mutual information (MI) about stimulus identity at each configuration of target and masker location. Here, MI serves as a measure of target discriminability for a given target-masker location. This discriminability is based on both the reliability of responses to a given target and dissimilarities in responses between targets. We expected that 1) MI would decrease when a masker is present versus when target-only trials, 2) MI would increase when the target is either presented by itself or spatially separated from the masking stimulus, and 3) MI would be reduced when the target location is ipsilateral due to the contralateral bias in the auditory hemisphere. To reduce the

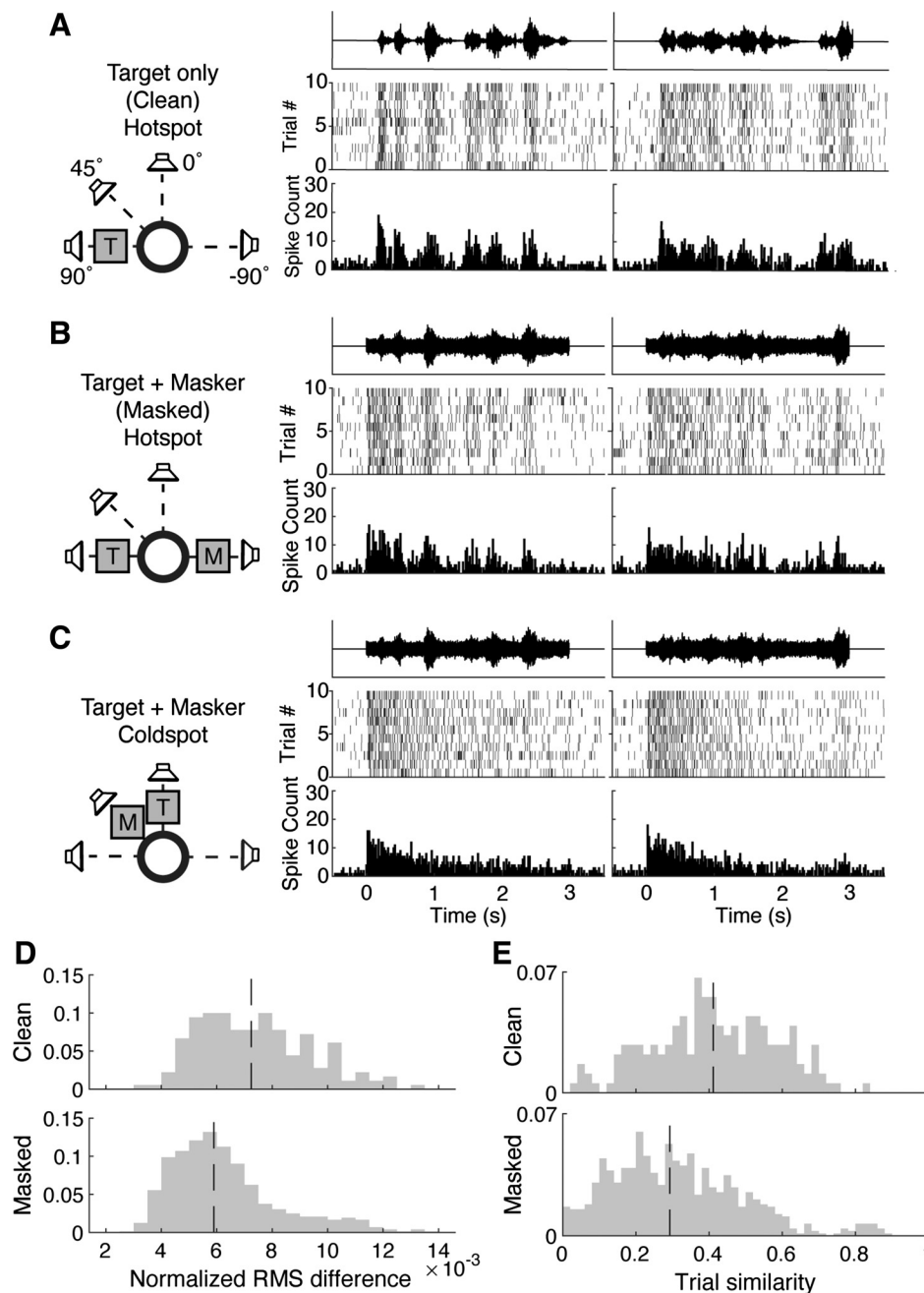


Figure 2. Cortical discrimination in a cocktail party paradigm in mouse auditory cortex (ACX). Portions of this figure are replicated from Nocon et al. (6), licensed under a Creative Commons license. **A:** stimulus configuration for clean trials originating at +90° azimuth and responses to both target stimuli. Auditory stimuli were presented from speakers at four locations. Target stimuli consisted of white noise modulated by human speech envelopes extracted from recordings of speech sentences (see METHODS). Responses during clean trials exhibit spike timing and rapid firing rate modulation that follow the amplitude envelope of both target stimuli, resulting in a neural discriminability value of 92%. All plotted peristimulus time histograms (PSTHs) have a bin length of 20 ms. **B:** stimulus configuration and responses for trials where targets (T) played at +90° and a competing masking stimulus (M) played at -90°. Masking stimuli consisted of unmodulated white noise with the same onset and offset times as target stimuli. In this configuration, spike timing and firing rate modulation follow both target stimuli, despite the presence of the competing masker, resulting in a neural discriminability value of 87%. **C:** stimulus configuration and responses for trials where targets played at 0°, and maskers played at +45°. In this configuration, spike timing and firing rate modulation do not follow either target stimulus, resulting in similar responses between target identities, resulting in a neural discriminability value of 60%. **D:** distribution of rate-normalized root-mean-square (RMS) difference between targets for all clean configurations ($n = 180$ configurations) and masked configurations ($n = 720$ configurations) across all 45 single units. Dashed lines represent median rate-normalized RMS differences for clean (median: 0.072) and masked (median: 0.059) trials. **E:** distributions of trial similarity for all clean trial configurations ($n = 180$ configurations) and masked trial configurations ($n = 720$ configurations) across all 45 single units. Dashed lines represent median trial similarities for clean (median: 0.4113) and masked (median: 0.2926) trials.

effect of time resolution on analysis and to avoid the under-sampling problems associated with classical estimators of MI, MI was estimated using spike train distances (12, 13, 31). We calculated spike train distances using the SPIKE-distance metric, which measures the difference in local firing rate and spike timing between trains without the need for an additional parameter (35). We explored two schemes of population coding: the summed population (SP, Fig. 3A), where spike trains from different neurons are pooled, and the labeled line (LL, Fig. 3B), where spike trains are concatenated in time to preserve unit identity. In both approaches, we used a forward search algorithm (10), where neurons are added to the population based on how well they complement the mutual information of the current set of neurons,

starting from single unit values (Fig. 3C). To control for the deletion of coincident spikes in the SP code, we deleted the same spikes in the LL code, such that only a single neuron within the subpopulation spikes at a given time. With this control, we found that an average of 10 coincident spikes is deleted when all 45 single units are pooled together (Supplemental Fig. S3), the effect of which is very small compared with the number of remaining spikes.

Finally, because population searches in the labeled line case exhibited plateaus in mutual information as neurons were added to the population (Fig. 3C), we estimated the minimum number of neurons needed to reach optimal MI. This minimum number was defined as the smallest subpopulation of neurons whose MI was above 90% of the maximum

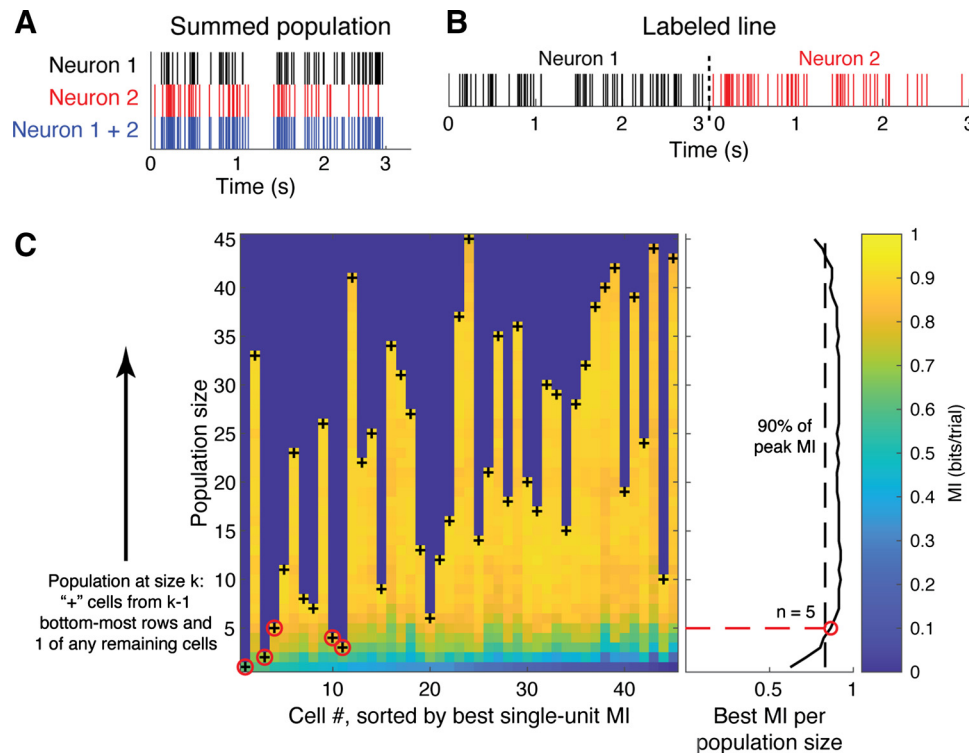


Figure 3. Population approaches and forward selection algorithm. *A*: Two spike trains in the summed population approach. Spike trains that occur during the same trial are pooled, resulting in a single neural response. *B*: example of labeled line approach. The same spike trains in *A* are concatenated in time to create a response whose length is the product of the trial time and the population size. *C*: the forward selection algorithm for determining the optimal subpopulation, starting with the mutual information (MI) values for all individual units. *Left*: matrix showing all MI values for each population size, with each element depicting the MI of one specific subpopulation. In each row, starting from single unit MI at the bottom row, “+” represents the unit that complements the current subpopulation of all prior “+” units, whether by yielding the largest increase or smallest decrease in MI. The forward selection search continues by adding each “+” (indicated by the upward arrow) until all neurons have been added. *Right*: line plot showing the best MI for a given population size, corresponding to the values shown in the matrix at each “+”. To estimate the minimum number of neurons needed to reach optimal MI, we used a threshold at 90% of the maximum MI across all populations (black dashed line). The population that first crosses this threshold was deemed as the optimal neural subpopulation K_{opt} with size n (red open circles in the matrix and line plot). Color bar shows the color scale for the mutual information values in the matrix.

value found across all values of n in the forward search for both population codes.

Upper Envelopes of Mutual Information

With our population search approach, we asked: what coding strategies improve upon single-unit target discriminability in the presence of competing stimuli? How does each coding strategy improve discriminability at certain configurations of target and masker location? We define the “upper envelope” as the maximal MI for each spatial configuration of target and masker location across all possible subpopulations of single units. We found that the summed population approach (Fig. 4*Ai*) yielded worse upper envelope MI values than that of the labeled line approach (Fig. 4*Aii*) in all configurations. In both cases, a diversity of neuron identities and layer locations was found to contribute to the upper envelope across all configurations, especially in configurations where the target was located at 0° .

When comparing how both approaches differed from the best single unit MI values (bottom-most dashed lines), we found that the labeled line code improved stimulus information coding across all configurations, whereas the single unit MI was optimal at some configurations for the summed population code. In some configurations of the summed population

code, we found that the forward search approach resulted in combinations of units where MI was lower than the best single unit value (Supplemental Fig. S4*A*), before additional units to the subpopulation resulted in improved MI. We attribute this nonmonotonic trend in MI to the destructive effect of the summed population code and the MI threshold criterion for the optimal subpopulation. In contrast, we find that MI in the labeled-line approach saturates at very small population sizes. As a result, adding more neurons to the optimal subpopulation results in very little to no change in MI until large values of n , at which we see a decrease (Supplemental Fig. S4*B*).

Number of Neurons in Optimal Subpopulations and Spatial Separation between Target and Masker

For the remaining analysis, we focused on all configurations where the target location is either contralaterally or centrally located relative to the recorded auditory cortex. When we compared the number of neurons that compose the MI upper envelope, we found that the labeled line code had larger subpopulation sizes than the summed population code. Indeed, the median number of neurons in K_{opt} from summed population searches was 1 (Fig. 4*Bi*), whereas the median number from labeled line searches was 2.5 (Fig.

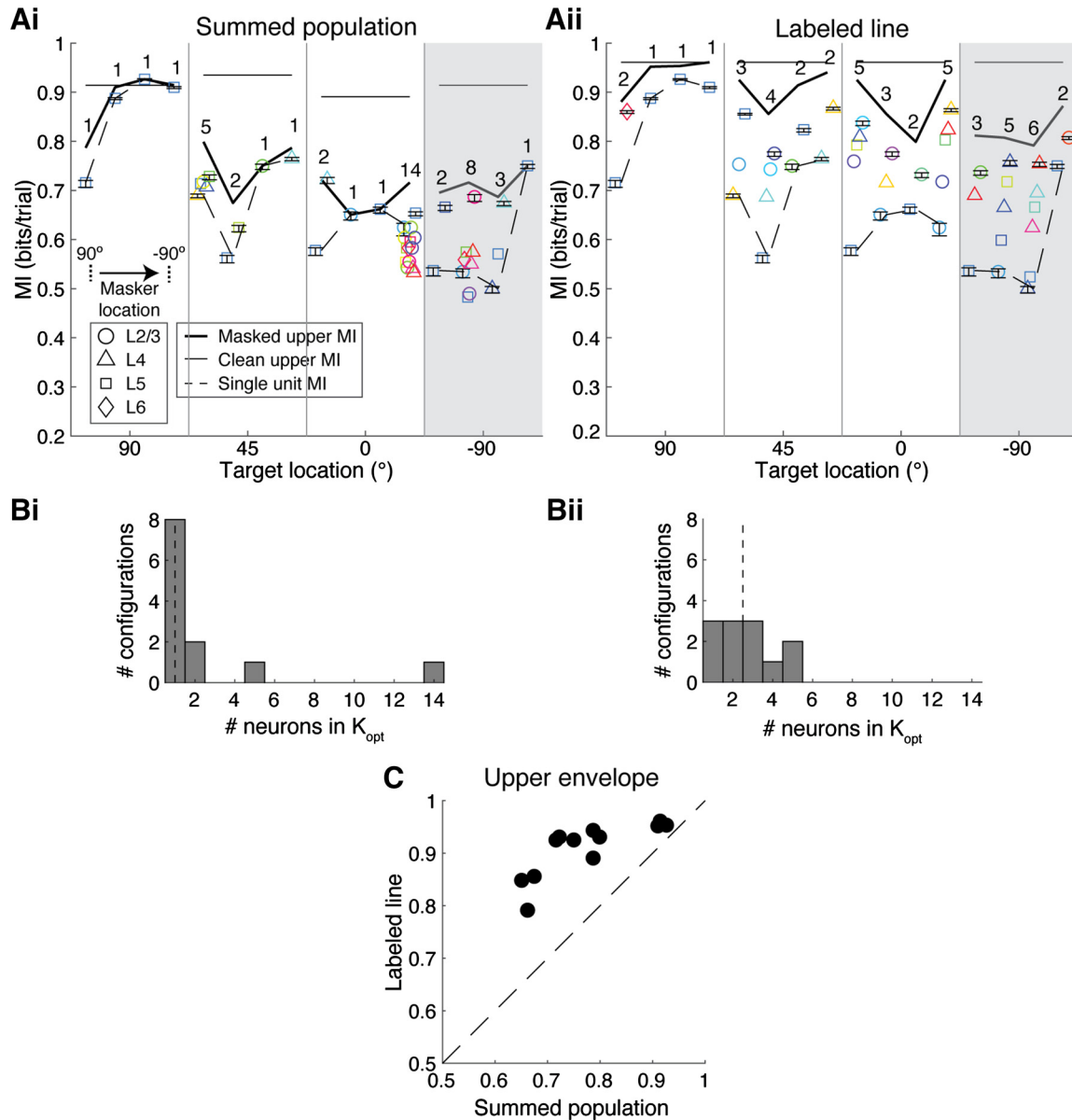
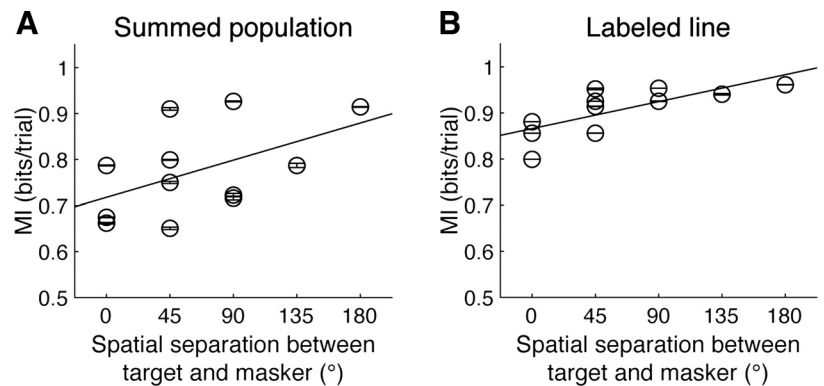


Figure 4. Upper envelopes of mutual information. **A:** upper envelope of mutual information (MI) for the summed population (SP, **Ai**) and labeled line (LL, **Aii**) approaches at all target-masker configurations. Shaded gray region denotes configurations where the target stimulus was ipsilateral (-90°) to the recorded auditory cortex hemisphere. Gray vertical lines separate the plot into sections based on the target location; within each section, the masker location is ordered from contralateral (90°) to ipsilateral (-90°), as shown by the arrow bounded with dashed lines in bottom-left of **Ai**. Within each target location section: thin black lines represent clean MI; bold black lines represent raw maximum MI at masked configurations; and dashed lines represent MI from the best single unit ($n = 1$ unit) at masked configurations, which is identical between approaches. Each configuration shows n , the number of neurons in K_{opt} . All units included in K_{opt} are shown as markers, with shape representing the cortical layer and color representing the identity of the unit and error bars representing $\pm SE$ from 100 iterations of leave-one-out resampling (see METHODS). For readability, error bars are only included for the best single unit and subpopulation MIs. For some configurations (SP configurations at -90° in SP, all configurations in LL), the highest incremental MI is lower than the upper envelope due to our estimations of K_{opt} . In addition, for some configurations in the SP approach, best single unit MI was higher than MI from combinations of single units (e.g., Target 0° , Masker -90° and Target -90° , Masker 45° , shown in Supplemental Fig. S2, for both SP and LL codes). One-way ANOVAs on the effect of layer in single-unit MI did not yield significance for the clean condition [$P(4,175) = 0.119$, $\eta^2 = 0.041$], but yielded a significant effect for the masked condition [$P(4,715) = 0.0198$, $\eta^2 = 0.016$]. However, none of the post hoc multiple comparisons yielded significant pairwise differences between layers. **B:** histograms of the number of neurons in K_{opt} for all nonipsilateral target configurations for the summed population (**Bi**) and labeled line (**Bii**) approaches, with the dashed lines representing the median number of neurons per configuration. **C:** scatter plot comparing the upper envelope of MI between both approaches at all nonipsilateral target configurations, with dashed line representing unity.

4Bii). In total, 13 unique units were required to optimally code MI across all optimal subpopulations with a SP code, whereas 16 unique units were required with an LL code. When directly comparing the two codes

the labeled line approach resulted in mutual information values that outperformed those from the summed population approach (Fig. 4C). These results, along with the finding that optimal subpopulations in the labeled line

Figure 5. Mutual information and spatial release from masking. Mutual information (MI) vs. spatial separation between target and masker with linear regression lines for both approaches, excluding all ipsilateral target configurations. Each data point represents the raw maximal MI for each masked configuration shown in Fig. 4 with error bars representing $\pm$ SE. There was no significant correlation between spatial release and stimulus mutual information for the summed population approach (A) ($r = 0.51$, $P = 0.091$), but this correlation yielded significance for the labeled line approach (B) ($r = 0.70$, $P = 0.0115$).



codes are larger than those in summed populations, are consistent with a previous study (7).

Finally, when we plotted the MIs versus the amount of spatial separation between target and masker, we found that MI increased with spatial release for both population approaches (Fig. 5). In both cases, there was a significant correlation between the spatial separation of target and masker and stimulus mutual information, which is consistent with findings on the effects of spatial unmasking of auditory stimuli in humans (36) and songbirds (5).

DISCUSSION

Bin-Less Estimation of Mutual Information

Previous studies on population coding of auditory stimuli have systematically varied bin sizes on spike times (8) or stimulus epochs (7), which affect calculations of mutual information based on spike trains. Here, we present an approach for the bin-less estimation of mutual information that involves the use of spike train distance, based on differences between spike trains (12). For this present study, we used a time scale-free spike distance metric where the distance between two trains is based on differences in spike timing and local rate (10, 35). The combination of a time scale-free distance metric and estimator decreases the effects that firing rate would have on mutual information results and avoids the need to determine time scales that reflect the neural population of interest (37). Although past studies have shown that the calculation of mutual information is robust against the choice of metric, including those with a time scale, these studies have also shown that while mutual information calculations demonstrate a mild dependence on the choice of time scale (13), these calculations have little dependence on the chosen metric, with the exception of firing rate that shows poor decoding accuracy (38).

Studies combining neurophysiological and psychophysical approaches have shown that sensory areas drive higher cortical regions, which integrate past and current information to form behavioral decisions (39). Similarly, ACx exhibits the ability to encode behavioral choices during tasks, even before prefrontal regions (40). These findings suggest a processing chain between ACx and upstream cortical regions that features both top-down and bottom-up connections. Our data were collected from recordings in passive conditions, which limited our ability to explore the ACx's role in behavioral tasks. Future studies can use our cocktail party-

like paradigm with attentive tasks and recordings from other regions to elucidate how behavioral states affect ACx responses. In such studies, our methods for calculating mutual information can be readily applied to compare single-unit responses from different areas. Here, we use this method to calculate target discriminability from responses from the same single unit or subpopulation, but prior studies demonstrate that this method can calculate the mutual information between two neurons with a common input (13). Our method could also be used to calculate mutual information about behavioral choices, for example, go left versus go right, in a decision-making task.

Comparisons between SP and LL Codes

In this study, we used two different population coding schemes to determine how stimulus information is optimally coded within mouse ACx. The key difference between these two approaches is whether neuron identity is maintained: in SP, responses are collapsed across time, resulting in a single response. Because of this, SP coding appears to have a destructive effect on the temporal features of individual neural responses, which we previously found to be especially important for high neural discrimination of dynamic stimuli (6). Indeed, for the configurations with contralateral target locations, the median optimal population size was 1 during summed population coding. In this approach, only a fraction of single units was required to span all possible combinations of target and masker locations. In contrast, individual neuron identities and temporal response features were kept intact in the LL code. This approach to coding yielded higher optimal MI values at all contralateral target configurations and improved upon the best MI from single units. Finally, we found that the values for MI for ipsilateral target configurations were consistently lower compared with contralateral target positions. This is likely due to the fact our recordings sampled units only from one hemisphere and the contralateral bias in ACx.

When comparing the upper envelopes from both population codes, we found that the optimal subpopulation for MI at 90°-target configurations consisted of a single unit. In masked configurations with the same target location, MI approached the ceiling of clean values. For other contralateral target locations, the optimal subpopulation size was typically larger, with the LL code providing significantly higher MI and reaching levels near the MI for clean targets. Thus, a LL code provided a highly noise-robust code compared with the SP, which was more susceptible to noise. Notably, for

certain target-masker configurations, optimal subpopulations in the LL code consisted of single units from multiple cortical layers. Although our analysis showed a significant effect of cortical layer on single unit MI, post hoc comparisons were unable to pinpoint significant differences between layers. We attribute this result to the low group size in certain layers, namely, L1 (1 unit) and L6 (2 units). To better analyze how layer affects single-unit MI, additional recordings with a better sampling of high-performing L1 and L6 units are required. Nevertheless, our findings suggest that a combination of cells from different layers facilitates high neural discriminability in the presence of competing spatial stimuli.

Despite our findings that discriminability in the labeled line code outperformed that of the summed population code, both approaches yielded a similar number of unique single units across all optimal subpopulations from each target-masker configuration. Thirteen out of the 45 analyzed single units were a member of at least one subpopulation in the SP code, while 16 of these single units were a member of at least one subpopulation in the LL code. Past studies have found that sparse coding within the auditory cortex maintains stimulus representations (26), even in the presence of background noise (41). Our implementation of the LL code concatenates individual responses along the time axis (8), which is analogous to a multidimensional vector in which each neuron's response contributes to a feature vector. Although it remains unknown if and how such a representation is implemented mechanistically in the brain, theoretical studies have proposed biologically plausible networks that can decode such multidimensional representations using working memory (42, 43) or balanced networks (44).

Spatial Release from Masking Increases Stimulus Mutual Information

Previous studies within cortical responses in songbirds have found that neural discriminability increases with spatial separation between target and masking stimuli. We found that spatial separation between target and masker significantly increased the MI for the LL code but not the SP code. These findings are consistent with similar results from behavioral experiments involving both speech and non-speech stimuli showing spatial release from masking (36, 45). Thus, the LL code better reflects such behavioral data.

Limitations of Study

One limitation of our analysis is our sampling of ACx. From our previously collected data, we identified a total of 137 single units across 12 mice, all recorded from right hemisphere. Here, we restricted our population to the 45 single units that showed at least one hotspot of high neural discriminability, and our analysis did not include resampling. In comparison, Downer et al. (8) used 278 single units across two subjects, and their analysis involved repeated sampling of subpopulations of 20 neurons each. In contrast, Ince et al. (7) used 49 neurons across three subjects and did not restrict their population to responsive units only for unbiased analysis. Despite the differences in population sizes and sampling methods, our analysis showed that labeled line code quickly approaches saturation values at around a few (~5) neurons, whereas summed population codes show the highest target

discriminability when the population is very small or close to 1, both of which agree with these previous studies.

For population coding analysis, neural populations were optimized using a forward search approach. We found that a brute force search consisting of all possible neuron populations did not supplement our analysis, as single-unit MIs were already for half of the configurations in the summed population code and labeled line-based MI approached ceiling values for some configurations with just one iteration of the forward search. Although a brute force analysis would have given us the true subset of neurons that compose the upper envelope, we found that the final step of estimating the minimum population size at each configuration provided the best tradeoff between accuracy and computational overhead. In addition, the population of single-unit neurons used for this analysis was restricted to those that showed at least one performance hotspot during complex scene analysis and thus did not include any nonencoding neurons. We were interested in determining if a population of multiple neurons could improve upon single-unit results, which have already been shown to encode the two target stimuli during clean trials where only the target was present.

In this study, we defined population code as the aggregation of activity from multiple neurons, an approach that has been used by other studies on population coding within auditory cortex (7, 8). One alternative approach to uncover population coding strategies for sensory stimuli involves linearly weighted sums of single-unit outputs. The results from such an approach, however, depend on the model used to simulate population readout, as well as its architecture and parameters. Although the intrinsic reliability of single units could serve as a basis for weights, our discriminability metric also depends on the difference in temporal structure between target responses. Furthermore, while our analysis does not weight responses from individual neurons, single units with the highest discriminability take precedence in the forward search approach, in concordance with neurons showing high selectivity having higher weights in the weighted readout method (46). Another approach involves estimating the latent variables that underlie neural activity in populations (47), which has been used to represent neural population dynamics during motor tasks (48, 49) and responses to stimulus offsets (50). We did not use this method within our study, as we do not have a sufficient number or length of trials required to effectively model these factors. With more data, this approach can be applied to quantify and visualize population dynamics during the cocktail party problem. Such methods are complementary to our approach of calculating mutual information. Together these approaches could better unravel population dynamics during different brain states, e.g., passive versus attentive or different areas, e.g., primary sensory cortical areas like A1 and prefrontal cortex.

Finally, mutual information was based on how spikes in a train contribute to the discriminability between two dynamic auditory stimuli. Because of this, we did not expect the summed population results to greatly improve upon single-unit results, which we have already found to encode stimulus identity via template-based classifiers (6). Previous studies on population coding have used multiple stimuli (8) or binned responses in time (7), whereas our calculations of

SPIKE-distance were based on differences between trains across the entire period of stimulus playback. Future studies on population coding during complex scene analysis could use multiple target stimuli or stimuli with both spectral and temporal differences to better determine how each approach optimizes target discriminability in a complex auditory scene.

DATA AVAILABILITY

Data will be made available upon reasonable request.

SUPPLEMENTAL DATA

Supplemental Table S1 and Supplemental Figs. S1–S4: <https://doi.org/10.5281/zenodo.8273646>.

ACKNOWLEDGMENTS

We thank Michael Economo, Alberto Cruz-Martín, and Monty Escabí for helpful discussions and comments on the manuscript.

GRANTS

The work in this manuscript was funded by National Science Foundation (NSF) Grant No. IIS-1835270 (awarded to K. Sen and X. Han), NIH/National Institute of Neurological Disorders and Stroke (NINDS) Grant No. 1R34NS11742-01 (awarded to K. Sen and X. Han), NIH/National Institute on Deafness and Other Communication Disorders (NIDCD) Grant No. 1T32DC013017-01A1 (awarded to J. C. Nocon), and NSF Grant No. 2319321 (awarded to K. Sen and H. Gritton).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

J.C.N., H.G., and K.S. conceived and designed research; J.C.N. and H.G. performed experiments; J.C.N., J.W., and C.H. analyzed data; J.C.N. and K.S. interpreted results of experiments; J.C.N. prepared figures; J.C.N. drafted manuscript; J.C.N., H.G., C.H., and K.S. edited and revised manuscript; X.H. and K.S. approved final version of manuscript.

REFERENCES

- Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12: 4745–4765, 1992. doi:10.1523/JNEUROSCI.12-12-04745.1992.
- Wang L, Narayan R, Graña G, Shamir M, Sen K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27: 582–589, 2007. doi:10.1523/JNEUROSCI.3699-06.2007.
- McDermott JH. The cocktail party problem. *Curr Biol* 19: R1024–R1027, 2009. doi:10.1016/j.cub.2009.09.005.
- Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K. Cortical interference effects in the cocktail party problem. *Nat Neurosci* 10: 1601–1607, 2007. doi:10.1038/nn2009.
- Maddox RK, Billimoria CP, Perrone BP, Shinn-Cunningham BG, Sen K. Competing sound sources reveal spatial effects in cortical processing. *PLoS Biol* 10: e1001319, 2012. doi:10.1371/journal.pbio.1001319.
- Nocon JC, Gritton HJ, James NM, Mount RA, Qu Z, Han X, Sen K. Parvalbumin neurons enhance temporal coding and reduce cortical noise in complex auditory scenes. *Commun Biol* 6: 751, 2023. doi:10.1038/s42003-023-05126-0.
- Ince RA, Panzeri S, Kayser C. Neural codes formed by small and temporally precise populations in auditory cortex. *J Neurosci* 33: 18277–18287, 2013. doi:10.1523/JNEUROSCI.2631-13.2013.
- Downer JD, Bigelow J, Runfeldt MJ, Malone BJ. Temporally precise population coding of dynamic sounds by auditory cortex. *J Neurophysiol* 126: 148–169, 2021. doi:10.1152/jn.00709.2020.
- Aronov D, Reich DS, Mechler F, Victor JD. Neural coding of spatial phase in V1 of the macaque monkey. *J Neurophysiol* 89: 3304–3327, 2003. doi:10.1152/jn.00826.2002.
- Satuvuori E, Kreuz T. Which spike train distance is most suitable for distinguishing rate and temporal coding? *J Neurosci Methods* 299: 22–33, 2018. doi:10.1016/j.jneumeth.2018.02.009.
- Satuvuori E, Mulansky M, Daffertshofer A, Kreuz T. Using spike train distances to identify the most discriminative neuronal subpopulation. *J Neurosci Methods* 308: 354–365, 2018. doi:10.1016/j.jneumeth.2018.09.008.
- Houghton C. Calculating mutual information for spike trains and other data with distances but no coordinates. *R Soc Open Sci* 2: 140391, 2015. doi:10.1098/rsos.140391.
- Houghton C. Calculating the mutual information between two spike trains. *Neural Comput* 31: 330–343, 2019. doi:10.1162/neco_a_01155.
- Gritton HJ, Nocon JC, James NM, Lowet E, Abdulkerim M, Sen K, Han X. Oscillatory activity in alpha/beta frequencies coordinates auditory and prefrontal cortices during extinction learning (Preprint). *bioRxiv*, 2020. doi:10.1101/2020.10.30.362962.
- James NM, Gritton HJ, Kopell N, Sen K, Han X. Muscarinic receptors regulate auditory and prefrontal cortical communication during auditory processing. *Neuropharmacology* 144: 155–171, 2019. doi:10.1016/j.neuropharm.2018.10.027.
- Rothauer EH, Chapman WD, Guttman N, Nordby KS, Silbiger HR, Urbanek GE, Weinstock M. I.E.E.E. recommended practice for speech quality measurements. *IEEE Trans Audio Electroacoust* 17: 225–246, 1969. doi:10.1109/TAU.1969.1162058.
- Hawley ML, Litovsky RY, Culling JF. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J Acoust Soc Am* 115: 833–843, 2004. doi:10.1121/1.1639908.
- Pachitariu M, Sridhar S, Stringer C. Solving the spike sorting problem with Kilosort (Preprint). *bioRxiv*, 2023. doi:10.1101/2023.01.07.523036.
- Rossant C, Kadir SN, Goodman DFM, Schulman J, Hunter MLD, Saleem AB, Grosmark A, Belluscio M, Denfield GH, Ecker AS, Tolia AS, Solomon S, Buzsaki G, Carandini M, Harris KD. Spike sorting for large, dense electrode arrays. *Nat Neurosci* 19: 634–641, 2016. doi:10.1038/nn.4268.
- Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish AD. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* 131: 1–11, 2005. doi:10.1016/j.neuroscience.2004.09.066.
- Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, Kepecs A. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498: 363–366, 2013. doi:10.1038/nature12176.
- Monaghan JJM, Garcia-Lazaro JA, McAlpine D, Schaette R. Hidden hearing loss impacts the neural representation of speech in background noise. *Curr Biol* 30: 4710–4721.e4, 2020. doi:10.1016/j.cub.2020.09.046.
- Jung F, Yanovsky Y, Brankack J, Tort ABL, Draguhn A. Respiratory entrainment of units in the mouse parietal cortex depends on vigilance state. *Pflugers Arch* 475: 65–76, 2023. doi:10.1007/s00424-022-02727-2.
- Li L-Y, Xiong XR, Ibrahim LA, Yuan W, Tao HW, Zhang LI. Differential receptive field properties of parvalbumin and somatostatin inhibitory neurons in mouse auditory cortex. *Cereb Cortex* 25: 1782–1791, 2015. doi:10.1093/cercor/bht417.
- Morrill RJ, Hasenstaub AR. Visual information present in infragranular layers of mouse auditory cortex. *J Neurosci* 38: 2854–2862, 2018. doi:10.1523/JNEUROSCI.3102-17.2018.
- Hromádka T, Deweese MR, Zador AM. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6: e16, 2008. doi:10.1371/journal.pbio.0060016.
- Liang F, Li H, Chou X-L, Zhou M, Zhang NK, Xiao Z, Zhang KK, Tao HW, Zhang LI. Sparse representation in awake auditory

- cortex: cell-type dependence, synaptic mechanisms, developmental emergence, and modulation. *Cereb Cortex* 29: 3796–3812, 2019. doi:10.1093/cercor/bhy260.
28. Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435: 341–346, 2005. doi:10.1038/nature03565.
29. Hoglen NEG, Larimer P, Phillips EAK, Malone BJ, Hasenstaub AR. Amplitude modulation coding in awake mice and squirrel monkeys. *J Neurophysiol* 119: 1753–1766, 2018. doi:10.1152/jn.00101.2017.
30. Kozachenko LF, Leonenko NN. Sample estimate of the entropy of a random vector. *Problems Inform Transmission* 23: 95–101, 1987.
31. Kraskov A, Stogbauer H, Grassberger P. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 69: 066138, 2004 [Erratum in *Phys Rev E Stat Nonlin Soft Matter Phys* 83: 019903, 2011]. doi:10.1103/PhysRevE.69.066138.
32. Witter J, Houghton C. A note on the unbiased estimation of mutual information (Preprint). *arXiv* 2105.08682, 2021. doi:10.48550/arXiv.2105.08682.
33. Parker AJ, Newsome WT. Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21: 227–277, 1998. doi:10.1146/annurev.neuro.21.1.227.
34. Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K. Invariance and sensitivity to intensity in neural discrimination of natural sounds. *J Neurosci* 28: 6304–6308, 2008. doi:10.1523/JNEUROSCI.0961-08.2008.
35. Kreuz T, Chicharro D, Houghton C, Andrzejak RG, Mormann F. Monitoring spike train synchrony. *J Neurophysiol* 109: 1457–1472, 2013. doi:10.1152/jn.00873.2012.
36. Best V, Ozmeral E, Gallun FJ, Sen K, Shinn-Cunningham BG. Spatial unmasking of birdsong in human listeners: energetic and informational factors. *J Acoust Soc Am* 118: 3766–3773, 2005. doi:10.1121/1.2130949.
37. Victor JD. Binless strategies for estimation of information from neural data. *Phys Rev E Stat Nonlin Soft Matter Phys* 66: 051903, 2002. doi:10.1103/PhysRevE.66.051903.
38. Houghton C, Victor JD. Measuring representational distances: the spike-train metrics approach. In: *Visual Population Codes: Toward a Common Multivariate Framework for Cell Recording and Functional Imaging*, edited by Kriegeskorte N, Kreiman G. Cambridge, MA: The MIT Press, 2011, p. 213–244.
39. Romo R, Salinas E. Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci* 4: 203–218, 2003. doi:10.1038/nrn1058.
40. Napoli JL, Camalier CR, Brown AL, Jacobs J, Mishkin MM, Averbach BB. Correlates of auditory decision-making in prefrontal, auditory, and basal lateral amygdala cortical areas. *J Neurosci* 41: 1301–1316, 2021. doi:10.1523/JNEUROSCI.2217-20.2020.
41. Schneider DM, Woolley SM. Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79: 141–152, 2013. doi:10.1016/j.neuron.2013.04.038.
42. Druckmann S, Chklovskii DB. Neuronal circuits underlying persistent representations despite time varying activity. *Curr Biol* 22: 2095–2103, 2012. doi:10.1016/j.cub.2012.08.058.
43. Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang XJ. Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci USA* 114: 394–399, 2017. doi:10.1073/pnas.1619449114.
44. Denève S, Machens CK. Efficient codes and balanced networks. *Nat Neurosci* 19: 375–382, 2016. doi:10.1038/nn.4243.
45. Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR. Free-field release from masking. *J Acoust Soc Am* 90: 1355–1370, 1991. doi:10.1121/1.401927.
46. Zavitz E, Price NSC. Weighting neurons by selectivity produces near-optimal population codes. *J Neurophysiol* 121: 1924–1937, 2019. doi:10.1152/jn.00504.2018.
47. Pandarinath C, Ames KC, Russo AA, Farshchian A, Miller LE, Dyer EL, Kao JC. Latent factors and dynamics in motor cortex and their application to brain-machine interfaces. *J Neurosci* 38: 9390–9401, 2018. doi:10.1523/JNEUROSCI.1669-18.2018.
48. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV. Neural population dynamics during reaching. *Nature* 487: 51–56, 2012. doi:10.1038/nature11129.
49. Cowley BR, Kaufman MT, Butler ZS, Churchland MM, Ryu SI, Shenoy KV, Yu BM. DataHigh: graphical user interface for visualizing and interacting with high-dimensional neural activity. *J Neural Eng* 10: 066012, 2013. doi:10.1088/1741-2560/10/6/066012.
50. Bondanelli G, Deneux T, Bathellier B, Ostojic S. Network dynamics underlying OFF responses in the auditory cortex. *eLife* 10: e53151, 2021. doi:10.7554/eLife.53151.