Systems/Circuits

# Neural Codes Formed by Small and Temporally Precise Populations in Auditory Cortex

**Robin A. A. Ince,**[1,2] **Stefano Panzeri,**[2,3] **and Christoph Kayser**[1,2,4]

[1]Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, [2]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom, [3]Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, 38068 Rovereto, Italy, and [4]Bernstein Centre for Computational Neuroscience, 72076 Tübingen, Germany

The encoding of sensory information by populations of cortical neurons forms the basis for perception but remains poorly understood. To understand the constraints of cortical population coding we analyzed neural responses to natural sounds recorded in auditory cortex of primates (*Macaca mulatta*). We estimated stimulus information while varying the composition and size of the considered population. Consistent with previous reports we found that when choosing subpopulations randomly from the recorded ensemble, the average population information increases steadily with population size. This scaling was explained by a model assuming that each neuron carried equal amounts of information, and that any overlap between the information carried by each neuron arises purely from random sampling within the stimulus space. However, when studying subpopulations selected to optimize information for each given population size, the scaling of information was strikingly different: a small fraction of temporally precise cells carried the vast majority of information. This scaling could be explained by an extended model, assuming that the amount of information carried by individual neurons was highly nonuniform, with few neurons carrying large amounts of information. Importantly, these optimal populations can be determined by a single biophysical marker—the neuron's encoding time scale—allowing their detection and readout within biologically realistic circuits. These results show that extrapolations of population information based on random ensembles may overestimate the population size required for stimulus encoding, and that sensory cortical circuits may process information using small but highly informative ensembles.

## Introduction

The encoding of sensory information by populations of cortical neurons forms the basis of behavior (Pouget et al., 2003; Averbeck et al., 2006; Graf et al., 2011). Considerable work has shown that information appears to be distributed over larger populations rather than being localized to individual neurons. Often, individual neurons do not provide sufficient evidence to support perceptual performance (Gold and Shadlen, 2007; Engineer et al., 2008; Safaai et al., 2013), suggesting that the activity of many cells must be considered to obtain a reliable representation within large stimulus spaces. Indeed, several studies probing the relation of population performance with population size reported that

population information grows steadily with population size. This led to the hypothesis that in sufficiently large ensembles each neuron contributes independently to a population code (Gawne et al., 1996; Rolls et al., 1997; Reich et al., 2001; Quiroga et al., 2007) and that large populations may encode very high amounts of information (Abbott et al., 1996).

However, such a distributed code necessitates the simultaneous monitoring of large populations to recover the available information (Gold and Shadlen, 2007). While distributed monitoring is compatible with the high convergence observed in cortical circuits, experiments show that neural activity is sparsely distributed across the ensemble (Greenberg et al., 2008; Crochet et al., 2011; Barth and Poulet, 2012) and that few afferents can be sufficient to drive entire circuits (Douglas and Martin, 2007; London et al., 2010). Hence, perception may be driven by small rather than large and concurrently active populations (Huber et al., 2008; Kwan and Dan, 2012).

When trying to reconcile the distributed information in cortical populations with response sparseness, we noted that previous work suggesting a highly distributed code relied mostly on randomly selected subpopulations from an existing dataset. For example, studies on the scaling of information with population size often considered the average performance of randomly assembled populations as at least one option (Abbott et al., 1996; Rolls et al., 1997; Narayanan et al., 2005; Quiroga et al., 2007). We hypothesized that such unselective averaging might give a distorted view of population coding by effectively creating a homogenously informative population that does not exist in reality,

hence masking the presence of small ensembles that could make privileged contributions to sensory evidence. To test this hypothesis we studied the encoding of natural sounds in primate auditory cortex. We found that small ensembles are sufficient to recover essentially all the information available from the entire recorded population, consistently for all cases considered. These "optimized" populations rely on temporally precise responses, rather than on many neurons, and can be identified by a simple biophysical marker, their encoding time scale. This shows that averaging over randomly assembled ensembles may overestimate the population size required to represent a specific amount of information and that highly informative population codes can consist of few privileged neurons.

## Materials and Methods

The data were obtained as part of a previous study (Kayser et al., 2009) and is analyzed here for a different question than previously. Briefly, recordings were obtained from the auditory cortex of adult male rhesus monkeys (*Macaca mulatta*). All procedures were approved by local authorities (Regierungspräsidium Tübingen), were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC), and were in concordance with the recommendations of the Weatherall report (2006) on the use of nonhuman primates in research. Before the experiments, a form-fitting headpost and recording chamber were implanted under aseptic surgical conditions and general anesthesia (Logothetis et al., 2010). As a prophylactic measure, antibiotics (Enrofloxacin, Baytril) and analgesics (Flunixin, Finadyne vet.) were administered for 3–5 d postoperatively. The animals were socially (group) housed in an enriched environment under daily veterinary supervision.

*Recording procedures, data preprocessing, and auditory stimuli.* As described previously (Kayser et al., 2009), neural responses were recorded from caudal auditory cortex (fields A1, CM, and CL) of three alert animals from multiple microelectrodes (1–6 MOhm impedance), high-pass filtered (4 Hz, digital two pole Butterworth filter), amplified (Alpha Omega system), and digitized at 20.83 kHz. Recordings were performed in a dark and anechoic booth while the animals were passively listening to the acoustic stimuli. Spike-sorted activity was extracted using commercial spike-sorting software (Plexon Offline Sorter) after high-pass filtering the raw signal at 500 Hz (third order Butterworth filter). For the present analysis we restricted analysis to units classified as "single units," using the following criteria: a signal-to-noise ratio of the spike waveform (peak amplitude divided by signal SD) of >8 and <2% of spikes with interspike intervals shorter than 2 ms. In total, we included 49 units recorded from 8 sessions. We did not classify units as responsive or informative about the stimuli using additional criteria but we simply included all units into the population analysis as described below to obtain an unbiased perspective.

Acoustic stimuli (average 65 dB SPL) were delivered from two calibrated free field speakers (JBL Professional) at 70 cm distance. The stimulus consisted of a continuous 52 s sequence of natural sounds that was presented many times (39–60 trials per unit). This stimulus sequence was created by concatenating 21 snippets of various naturalistic sounds, each 1–4 s long, without periods of silence in between (animal vocalizations, environmental sounds, conspecific vocalizations, and short samples of human speech). See Figure 1*A* for a spectral representation.

*Neural response properties.* For each unit we calculated the encoding time and the life-time sparseness. The encoding time (ET; Theunissen and Miller, 1995; Chen et al., 2012) was obtained from the normalized autocorrelation of the unit's peristimulus time histogram (PSTH, 1 ms resolution) by fitting a Gaussian to the autocorrelogram (least-squares; skipping the zero lag). The encoding time was defined as twice the SD of the fitted Gaussian. The life-time sparseness was defined using the following index (Vinje and Gallant, 2000):

$$S = 1 - \left[ \frac{\left( \sum_i r_i / n \right)^2}{\sum_i r_i / n} \right], \qquad (1)$$

where $r_i$ is the trial-averaged response in the $i$th bin (using 40 ms width) of the PSTH and $n$ is the total number of time bins.

*Definition of population codes.* A pseudo-population response was obtained by creating a vector from the responses of a group of units sampled in time bins $\Delta t$. The use of a pseudo-population was necessary as only a subset of units were recorded simultaneously (range $n = 3$–9). To create the pseudo-population we shuffled trials independently for each unit to remove potential trial-by-trial correlations. In a separate analysis we verified that such correlations had a very small impact on the present results (see below). As slightly different numbers of trials were available for the cells recorded in different experimental sessions, we used resampling (imputation) to avoid limiting the analysis to the smallest number of trials available. We set a required number of 58 trials per unit, chosen such that at most 5% of the total number of trials needed to be added. Resampling was done by random sampling with replacement the required number of additional trials from the existing trials for each unit.

To calculate the stimulus information carried by these codes about a set of natural sounds we defined "stimuli" as segments of length $T$ sampled within the entire continuous stimulus sequence (Kayser et al., 2009, 2012). The length $T$ and the binning $\Delta t$ were varied and for each unit the response vector had dimensionality $M$ (with $T = M \times \Delta t$). In general, the responses of different units can be combined by either preserving the identity of each unit, a so-called "labeled line," or the responses can be averaged, a so-called "pooled" code (Perkel and Bullock, 1968; Reich et al., 2001). For a population of $N$ neurons this results in a matrix containing $M \times N$ responses for the labeled line code, with each element containing the response of the respective unit in each time bin. For the pooled code the responses are summed within each bin across neurons resulting in an $M \times 1$ vector. Formally, these codes are defined as $r = (r_1^1, r_1^2, \ldots, r_1^M, r_2^1, r_2^2, \ldots, r_N^M)$ for the labeled line and $r = (r_{pop}^1, \ldots, r_{pop}^M)$ for the pooled, with $r_i^k$ denoting the number of spikes emitted by unit $i$ in time bin $k$, and $r_{pop}^k = \text{Sum}_i(r_i^k)$.

The temporal binning $\Delta t$ was varied systematically (Fig. 6). As default we used 40 ms, a value that was chosen based on the encoding time scale of individual units and previous literature (Lu and Wang, 2004; Schnupp et al., 2006; Engineer et al., 2008; Kayser et al., 2010).

*Decoding and information calculation procedures.* We used a decoding procedure to estimate the mutual information between the population code and a set of stimuli. Specifically, we selected $N_s$ non-overlapping stimulus epochs of length $T$ from the continuous stimulus sequence, resulting in a stimulus ensemble $S = (S_1, \ldots, S_{N_s})$ that reflects a subset of sound tokens occurring within the experimentally presented sound sequence. For the main analysis we chose $N_s = 20$ and $T = 120$ ms but we varied these parameters systematically to ensure the general validity of results (Fig. 6). Population codes were defined for each epoch and trial as described above. A decoder was used to predict the stimulus associated with single trial responses using a leave-one-out cross-validation procedure. For each "test" trial the decoder was trained on the stimulus set consisting of all trials except this test trial and was used to predict the respective stimulus. Prediction performance was recorded in a confusion matrix, $Q(S_d|S_i)$, which contains the probability that a presented stimulus epoch $S_i$ is decoded as epoch $S_d$. The overall performance for each code was then measured by calculating the information in the confusion matrix $I(S_d;S_i)$ (Nelken and Chechik, 2007; Quian Quiroga and Panzeri, 2009; Kayser et al., 2010).

$$I(S; D) = \sum_{d,s} P(S_i) Q(S_d|S_i) \log_2 \frac{Q(S_d|S_i)}{Q(S_d)}. \qquad (2)$$

Where in the above equation $Q(S_d)$ is the marginal probability of the decoded stimulus. Information values were corrected for limited sampling bias using the Panzeri-Treves method (Panzeri and Treves, 1996; Magri et al., 2009).

We compared different decoding algorithms to ensure the validity of results. We focused on algorithms that permit an efficient implementation of the cross validation procedure (Hastie et al., 2009) and we considered a family of classifiers that use a generative multivariate normal model. This included diagonal and full versions of quadratic and linear discriminant classifiers, which estimate covariance matrices for each
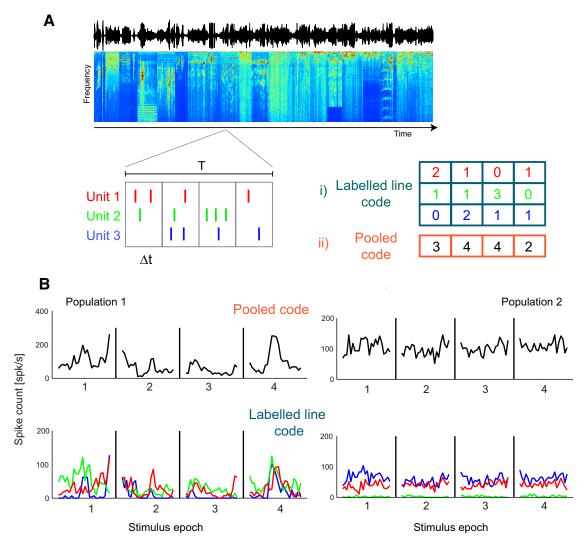
**Figure 1.** Illustration of population codes and example data. *A*, Time-frequency representation of the acoustic stimulus sequence (52 s duration) and schematic of a stimulus epoch (*T*) selected from this. Illustration of the responses of three hypothetical units divided into time bins (Δ*t*). The responses of different units can be combined by either preserving each unit's identity, the labeled-line code (blue), or by summing the responses within each time bin, the pooled code (orange). *B*, Example data showing actual responses of two populations consisting of *N* = 3 units during four stimulus epochs (Δ*t* = 5 ms, *T* = 120 ms). Information values were 1.23/2.0 bits (pooled/labeled line) for the informative example on the left and 0.13/0.22 bits for the example on the right. Lines indicate the PSTH for either the population (pooled code) or each individual unit (labeled line code).

stimulus, or common to all, and we tested Naive Bayes decoders, which assume that response features are independent and estimate the most likely stimulus using Bayes theorem (Kayser et al., 2012). While we found generally very comparable results, the diagonal linear classifier exhibited best overall performance. Specifically, this decoder assumes the response features to be independent and can be formalized as follows: let $m_{is}$ be the mean of each response feature $i$ to each stimulus $s$ (where a response feature is the spike count within a specific time bin for a particular unit in the case of the labeled line code, or summed over the population for the pooled code, see previous section), and let $\sigma_i^2$ be the corresponding variance of element $i$, both calculated by excluding the current test trial. The test trial $\mathbf{r} = (r_1, \ldots, r_n)$ is then classified as the stimulus that maximizes the following discriminant function:

$$g_s(\mathbf{r}) = -\frac{1}{2}\sum_i \left(\frac{r_i - m_{is}}{\sigma_i^2}\right), \qquad (3)$$

which is equivalent to minimizing the normalized Euclidean distance between the test trial and the stimulus response clusters assuming a uniform prior over stimuli (Alpaydin, 2010). To avoid numerical problems due to ill-conditioned covariance matrices, we added a small random jitter (normally distributed with SD 0.01) to the discrete spike count

responses independently for each trial and bin. We verified that this addition of jitter gave similar results to the more computationally intensive use of the matrix pseudo-inverse.

For each unit, we defined the single unit information (e.g., used in Eq. 6 or Fig. 4) as the average information provided by that unit across the 100 randomly sampled stimulus ensembles of size $N = 20$.

*Definition of optimized subpopulations.* The decoding procedure was applied to different subpopulations sampled from the total available neural ensemble and separately to pooled and labeled line codes. Our main goal was to compare randomly chosen populations to populations selected to maximize the mutual information conveyed about the stimulus. We therefore determined optimized, maximally informative populations for a given stimulus ensemble (of size $N_s$) using a forward selection procedure (Hastie et al., 2009). We built a population of size $N+1$ by adding to the already existing population of size $N$ the unit that provided the largest information increment. Specifically, for each stimulus ensemble $S$ we started by selecting the most informative unit. We then calculated the information for all populations of size $N = 2$ by adding each of the remaining units in turn and determining the pair with highest information. This most informative pair was defined as optimal population for size $N = 2$ and we proceeded by testing each of the remaining units

**Table 1. Comparison of information values provided by 'optimized' populations derived using exhaustive search**

| Population size $N$ | Labeled line | Pooled |
|---|---|---|
| 5 | $0.998 \pm 0.0023$ | $0.967 \pm 0.031$ |
| 10 | $0.994 \pm 0.0023$ | $0.954 \pm 0.044$ |
| 15 | $0.991 \pm 0.0026$ | $0.954 \pm 0.034$ |
| 20 | $0.990 \pm 0.0020$ | $0.972 \pm 0.014$ |

The table lists the mean ratio (mean $\pm$ SD over 10 populations) of the information conveyed by the optimized population determined by the forward selection algorithm used for the main analysis and the true optimal population obtained through an exhaustive search probing all possible populations. The forward selection procedure yields populations providing on average at least 95% of the maximal possible information for each $N$.

when added to form a population of size $N = 3$. This procedure was repeated until all units were included, resulting in a ranking of units in terms of which step (population size) they were added to the cumulative optimized population. When we varied parameters such as the size of the stimulus ensemble ($N_s$) or stimulus epoch duration ($T$), we independently repeated the forward selection procedure. As comparison to this optimized population, we calculated the information provided by randomly selected subpopulations of a given size. These were obtained by sampling units at random from the full set, without replacement, and for each stimulus ensemble $S$ we averaged the performance of 100 random populations of a given size to obtain one value for random populations associated with each stimulus ensemble.

It is important to note that the forward selection procedure described above is a local search heuristic and is not guaranteed to obtain the global optimum population for each population size. While an exhaustive search was not computationally feasible for the full range of population sizes considered here, we performed the following control to assure that our algorithm provides results that are in close agreement with the true optimal solution. Specifically, we compared the information in populations obtained with the forward selection algorithm to the true optimal populations obtained from an exhaustive brute force search. To facilitate the exhaustive search we had to limit the total population from which subpopulations were sampled. Rather than using all 49 units, we used randomly selected test populations of total size 25 and sampled subpopulations of size $N = 5, 10, 15,$ and $20$ neurons from these 25 cells. The information was calculated using the methods described above ($N_s = 20$; $T = 120$ ms; $\Delta t = 40$ ms) with a set of 20 stimulus ensembles, and the process was repeated 10 times using different test populations of size 25. Results of this calculation (shown in Table 1) revealed that the forward selection heuristic provides a good approximation to the true optimal population and yields populations that provide >95% (>99% for labeled lines) of the maximal attainable value. Note that the forward selection provides an enormous computational saving over the brute force approach. For example, for populations of size 49 as considered here, the brute force approach would require on the order of $10^{14}$ information calculations (across all population sizes for a single stimulus ensemble), while the forward selection procedure requires only 1224.

*Theoretical scaling of information in populations of units with random information overlaps.* The scaling of information with population size can be analytically treated under specific assumptions about the population. Previous work used this to derive estimates of the dependence of information on population size in partly redundant neurons (Gawne and Richmond, 1993) or to approximate the scaling of information in homogenous populations derived from random averaging (Rolls et al., 1997). The basic assumption is that that each neuron carries an amount of information that is drawn (from the total information space constituted by the given stimulus ensemble) independently from the information provided by any other neuron. Under this assumption, any "overlap" between the information provided by each neuron arises purely from the random sampling within the limited entropy in a fixed stimulus set. If one considers a homogenous population or uses this model to approximate a population homogenized by averaging over many randomly assembled populations, a single parameter is sufficient to describe the scaling of information with population size. Within this model, the information $I_M$ provided by $M$ neurons about $N_s$ equally probable stimuli is given by (Gawne et al., 1996; Rolls et al., 1997):

$$I_M = (1 - \Phi^M)\log_2(N_s). \tag{4}$$

Here, $\log_2(N_s)$ corresponds to the total information needed to perfectly discriminate all stimuli and $\Phi$ denotes the fraction of this information that, on average over the considered population, is still missing after observing a single neuron response. Hence, $\Phi = 1 - I/\log_2(N_s)$, with I being either the average of the single neuron information (when considering a homogenized population) or the single neuron information of a truly homogenous population. Equation 4 can be derived as follows: by definition each unit provides a fraction $1-\Phi$ of the $\log_2 N_s$ bits required to discriminate all $N_s$ stimuli. Assuming that the information overlap between neuron is random and due to the finite stimulus set, a fraction $\Phi^2$ is missing (on average) when considering two units, and so on. The fraction of information needed for stimulus discrimination still missing when considering $M$ neurons is thus on average $\Phi^M$, and Equation 4 follows. To fit Equation 4 to the performance of the average random population, we estimated the parameter $\Phi$ by fitting the above expression to the average information provided by randomly selected populations of size $M$ using the method of nonlinear least squares.

This model can be extended by allowing the population of neurons to have inhomogeneous single unit information values, while still maintaining the assumption that any overlap between the information of different units arises purely from random sampling within the stimulus entropy. Under this less restrictive assumption, the formula describing the scaling of information with population size becomes:

$$I_M = (1 - \textstyle\prod_{i=1}^{M}\Phi_i)\log_2(N_s), \tag{5}$$

where, $\Phi_i = 1 - I_i/\log_2(N_s)$, and $I_i$ denotes the information provided by neuron $i$. We fitted this model to data by first parameterizing the distribution of single unit information (see Fig. 3B) using an exponential distribution in which the information conveyed by the $i$th most informative neuron is given by $I_i = ae^{bi}$, where $a$ and $b$ are the two parameters of the model. The fraction of information missing after observing neuron $i$ then becomes the following:

$$\Phi_i = 1 - \frac{ae^{bi}}{\log_2(N_s)}. \tag{6}$$

We fit this inhomogeneous-information random-overlap model to both the average information provided by randomly selected populations, and the average information conveyed by the optimized populations. We used the method of nonlinear least squares to determine the values of the two parameters ($a$, $b$).

*Population statistics.* We computed two indices typically used to study population codes: population sparseness and population dispersal (Willmore and Tolhurst, 2001; Weliky et al., 2003). Population sparseness was calculated using the same expression as for lifetime sparseness above (Eq. 1), but with $r$ representing the population response and $r_i$ representing the trial-averaged response of the $i$th unit. Sparseness was calculated by dividing the full acoustic stimulus sequence into non-overlapping 120 ms windows and sparseness values were averaged across these windows. Dispersal indexes the relative spread of response variability across all neurons, whereby response variability refers to the response variance across stimuli. In a highly dispersed code many neurons respond differently to different stimuli whereas in a little dispersed code only few neurons exhibit large response variations across the stimulus ensemble. Dispersal was calculated as follows. The variance of the mean response (PSTH) across stimulus epochs was calculated for each unit in the population. These variances were normalized by dividing by the maximum variance over all units. The population dispersal statistic was obtained by taking the sum over units of the normalized variance, divided by the number units ($N = 20$). In this way the value is normalized for population size, providing a unit free quantity that can be compared across studies.

*Effect of noise correlations.* Our analysis is based on pseudo populations constructed from partly simultaneously and partly independently recorded units and does not include the effect of trial-by-trial ("noise") correlations (Panzeri et al., 1999; Averbeck et al., 2006). Although such correlations are generally weak (Schneidman et al., 2006; Ecker et al.,
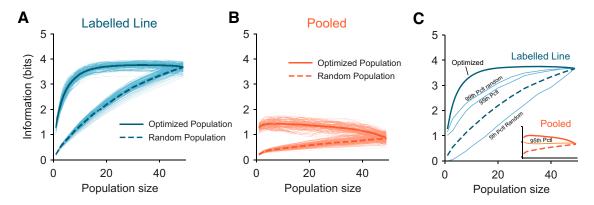
**Figure 2.** Stimulus information for optimized and random populations. *A*, *B*, Stimulus information for optimized populations obtained using forward selection and random populations as a function of population size. Thin lines indicate performance for individual stimulus ensembles (size $N_s = 20$) and thick lines indicate the mean across ensembles (other parameters $T = 120$ ms; $\Delta t = 40$ ms). *C*, Distribution of the stimulus information across random populations for the labeled line code (main) and pooled code (insert). For each random population, information was averaged across 100 stimulus ensembles ($N_s = 20$; $T = 120$ ms; $\Delta t = 40$ ms). The dashed line indicates the mean across random populations, thin lines the fifth, 95th, and 99th percentiles (over 200 random populations). Information in the optimized population (averaged across individual optimal populations for each of 100 stimulus ensembles) is shown as thick solid line.

2010), they can either decrease or increase the population information (Panzeri et al., 1999; Averbeck et al., 2006). We confirmed that the omission of noise correlation in the present analysis does not have an impact on the scaling of information, at least for small populations ($M \leq 6$). To this end we considered sets of simultaneously recorded units ($M = 3, 4, 5, 6$; with 245, 287, 231, 127 populations, respectively) and computed the information using either the actual population responses (i.e., including noise correlations) or the pseudo-population (i.e., the trial-shuffled data). The difference between these was very small ($<2\%$ for each population size) and did not increase with population size (one-way ANOVA, $p = 0.56$). This suggests that the use of pseudo-populations in the present case gives an accurate estimate of the information provided by simultaneously recorded neurons, at least for population sizes in the range of up to few tens of units. This is the relevant range for this study, as described in the following.

## Results

We considered the encoding of individual sound tokens occurring within a continuous stream of natural sounds (mostly environmental and animal sounds). For analysis we mimicked this by dividing the continuous sound sequence presented during the experiment (52 s duration) into epochs of length $T$ (Fig. 1A). For each unit we sampled the spike count in subsequent time bins of width $\Delta t$ within the stimulus epoch $T$, ranging from short bins emphasizing high response precision to a coarse temporal code (using $\Delta t = 40$ ms and $T = 120$ ms as default, unless stated otherwise). The stimulus discriminability afforded by each code was quantified by randomly sampling ensembles of stimulus epochs from the entire sound sequence (default $N_s = 20$) and using these as stimuli for a decoding analysis. To provide generic insights we considered two alternatives for creating a population code from a given set of responses and we report results for both: the labeled line, which preserves the identity of each unit, and the pooled code, which averages the responses across units (Perkel and Bullock, 1968; Fig. 1A, right). Example data for each code are shown in Figure 1B, with the example on the left/right showing a more/less informative population responses.

### Information in random and optimized populations

Our main focus was the selection of neurons included in a population and its impact on the information provided by this population. Specifically, we wondered whether there are small subsets of neurons that make privileged contributions to a population code and whose presence is masked by methods relying

on unselective and random averaging of subpopulations, i.e., methods that consider the average behavior of randomly assorted subpopulations. To this end we compared the information carried by subpopulations consisting of units randomly selected from the entire recorded ensemble with that carried by populations optimized for providing information about the current stimulus set. These maximally informative populations (termed "optimized" in the following) were determined through forward selection, in which, starting from the single most informative unit, the optimized population was expanded by iteratively adding that unit providing the largest information increment. For a given population size $N$ this provides a subpopulation from the total recoded ensemble whose information by construction closely approximates the highest achievable values for subpopulations of that size and for the given stimulus set. We then compared the performance of optimized and random populations for a range of parameters. While this forward selection is a heuristic algorithm and is not guaranteed to find the true optimal population, we verified its performance against a brute force search (see Materials and Methods) for reduced population sizes. This confirmed that the forward selection found solutions that provided on average 95% or more of the information contained in the true optimal solution (Table 1).

Figure 2 displays the information provided by randomly selected and optimized populations over many ensembles of stimulus epochs ($N_s = 20$), for each of which we constructed the optimized population. For random populations, information increases monotonically with population size $N$ for both labeled line (Fig. 2A) and pooled (Fig. 2B) codes. However, when considering the optimized populations, the scaling of information with population size was markedly different. Relatively few cells were sufficient to achieve the highest obtainable information. For the labeled line code, information increased steeply with population size for $N < 20$, and on average across stimulus ensembles $14.96 \pm 2.3$ units (mean $\pm$ SEM across 100 stimulus ensembles) were sufficient to obtain 95% of the maximally attainable information. Note that this maximally attainable information is the same for optimized and random populations and for the latter is attained when including all units. The selection of an appropriate set of ~30% of units hence provides an amount of stimulus information that would require a much larger number of units when selected at random. For the pooled code, the maximal ob-

tainable information differed between optimized and random populations due to the possible destructive interference when averaging the responses of many neurons. Performance of the optimized population peaked for $N = 9.63 \pm 6.71$ units and the inclusion of additional units reduced information. Notably, with the pooled code, the information conveyed by the full population was considerably lower than that obtained with smaller optimized populations ($57.1 \pm 6.5\%$ relative to the optimal population). This illustrates that the optimized selection of a population can provide larger amounts of information and with fewer units than achievable using randomly sampled populations.

When compared with the performance of many individual random populations, the average performance of the optimized populations fell well above the 99th percentile for essentially all population sizes and both kinds of code (Fig. 2C). The optimized population thereby provides a large and significant information gain compared with the average random population. Hence, when considering subsets of neurons from a given ensemble, one should not rely on randomly selected subpopulations (or averages thereof) as this may bias results, for example by considerably underestimating the available information.

Note that the above analyses are based on pseudo-populations. However, we verified that within the range of population sizes that we could test, the difference in information between pseudo- and simultaneously recorded populations was very small and did not increase with population size (see Materials and Methods). This suggests that these results suffer only marginally from the use of pseudo-populations.

## Theoretical information scaling of independent populations in a labeled line code
In the labeled line code the increase in information with population size in random populations was steady, yet less than linear. If each neuron in a population carried independent information one would expect a linear increase of information (Schneidman et al., 2003). However, the use of finite stimulus ensembles can introduce an overlap between the information carried by independent neurons simply as a result of limited stimulus entropy (Gawne et al., 1996; Rolls et al., 1997). We tested whether the scaling of information in our data can be explained by models of neurons that carry independent information apart from random redundancies due to the finite stimulus entropy. Specifically, we considered two such models, an "homogenized" model (Eq. 4) assuming that the single unit information had a constant value across units (equal to the ensemble average of single unit information), and a "heterogeneous" model that allowed units to have variable amounts of information, with a distribution of information values matching the experimentally observed exponential distribution (Eq. 5).
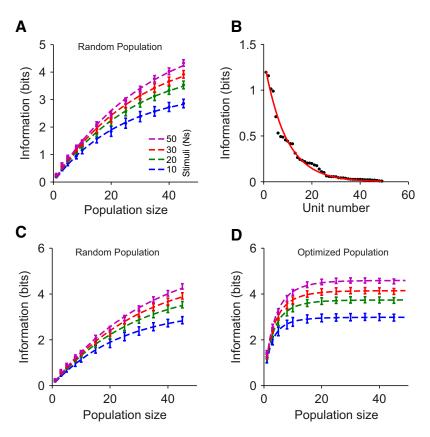


**Figure 3.** Scaling of population information matches prediction of theoretical models. **A**, Information in random populations using the labeled-line code for stimulus ensembles of different size (mean $\pm$ SD over 100 stimulus ensembles). Dashed lines indicate model fits assuming independent and equal contributions of each unit (Eq. 4; compare Materials and Methods). **B**, The distribution of single neuron information values is fit by a two parameter exponential distribution. **C**, Information in random populations using the labeled-line code for stimulus ensembles of different size (mean $\pm$ SD over 100 stimulus ensembles). Dashed lines indicate model fits assuming different information values for each neuron, following an exponential distribution as in **B**. (Eq. 5; compare Materials and Methods). **D**, Information in optimized populations using the labeled-line code for stimulus ensembles of different size (mean $\pm$ SD over 100 stimulus ensembles). Dashed lines indicate model fits assuming different information values for each neuron (Eq. 5; compare Materials and Methods).

Figure 3A shows that the scaling of information in random populations (with their performance averaged across 100 stimulus ensembles) can be well reproduced by the homogenized single-parameter information model (Eq. 4). We confirmed this for stimulus ensembles of variable size ($N_s = 10–50$ stimulus epochs) and in each case the average information in random populations scaled according to the prediction of the homogeneous model ($R^2 > 0.99$ for all $N_s$). The performance of the optimized populations, in contrast, could not be accounted for by this homogenized random-overlap model. However, we found that the extended model allowing for the variable information contributed by each unit reproduced the scaling for the optimized populations well. For the present data, we found that the distribution of single unit information across all recorded neurons was well fit by an exponential distribution (Fig. 3B; $R^2 = 0.98$), conforming to the general idea of sparsely distributed single neuron responses (Baddeley et al., 1997; Lehky et al., 2011; Willmore et al., 2011). The extended model fit well both the average information in random populations (Fig. 3C; all $R^2 > 0.99$), and the information in the optimized populations information (Fig. 3D; all $R^2 > 0.99$). This shows that the growth of information in multineuron populations can be well accounted for by models that assume (1) the near-independent contributions of neurons that carry highly unequal amounts of individual stimulus information, and (2) assume that any information over-
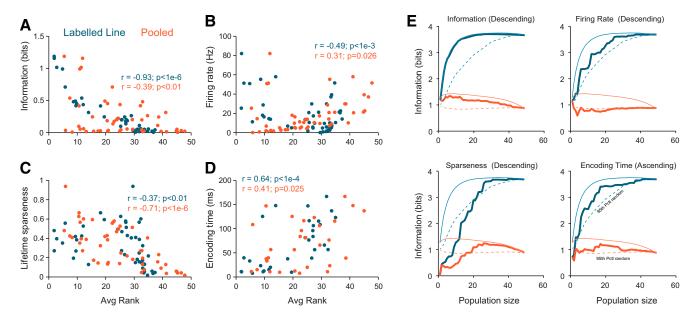
**Figure 4.** Response properties predictive of neurons contributing to optimal populations. ***A–D***, Scatter plots displaying the average rank of each neuron within optimal populations versus the single neuron information (***A***), firing rate (***B***), lifetime sparseness (***C***), and encoding time (***D***) associated with that unit. The average rank was calculated across 100 sets of stimulus epochs ($N_s = 20$). ***E***, Average stimulus information (mean over 100 stimulus ensembles, $N_s = 20$) for populations constructed based on the systematic addition of neurons according to their rank order for different response properties (thick lines). For comparison, performance of the optimized (thin lines) and the upper 95% percentiles of random populations (dashed) are shown.

lap between units is only due to the finite amount of information in the experimental stimulus set.

## Properties of units included in highly informative populations

Given that the ensemble of neurons sampled to constitute the population codes proved heterogeneous, at least with regard to sensory encoding, we characterized those units contributing to the optimized populations in further detail. We compared the unit's average rank within the optimized populations (computed over 100 different stimulus ensembles) to specific neural properties: their individual stimulus information, mean firing rate, lifetime sparseness (Vinje and Gallant, 2000), and encoding time scale (Theunissen and Miller, 1995). The results reveal (Fig. 4A–D) that units with low rank in labeled lines individually carried high information ($r = -0.93$), had high firing rates ($r = -0.49$), and a short encoding time ($r = 0.64$; all $p < 0.01$). Units with low rank in a pooled code had high life-time-sparseness ($r = -0.71$), lower firing rates ($r = 0.31$), and short encoding times ($r = 0.41$; all at least $p < 0.05$). This highlights key differences between labeled line and pooled codes. A pooled code can suffer from including units with high firing rate but is likely to benefit from units that are sparsely active, while a labeled line code benefits most from including units that are themselves very informative.

Importantly, these results also suggest that by selecting units based on one of these response properties, rather than performing the artificial optimization procedure based on information, one may be able to obtain a highly informative subpopulation. We tested this directly by building subpopulations by adding units according to their response properties and comparing them to random and information-theoretic optimized populations. Practically, for populations based on single unit information and encoding time we added units in ascending order of the respective property and for populations based on firing rate or sparseness in descending order. The results (Fig. 4E) show that selecting populations based on the most informative single units provides

a close approximation to the optimized populations. However, the single unit information may be difficult to assess for a downstream decoder within biological circuits, as it constitutes an artificial construct. However, the other parameters have a more direct biophysical interpretation and could in principle be extracted by cortical circuits. Of those features, selecting units based on their encoding time scale provided the best approximation to the optimized population; with firing rate being similarly efficient for the labeled line code. For both codes, selection based on encoding time provided a population that fell well outside the 95% percentile of random populations and provided only a small information loss compared with the optimized population. The mean information ratio (averaged across population sizes) between the feature-based and information-optimized populations were 0.99/0.88 (labeled line/pooled) for single unit information, 0.88/0.79 for encoding time, 0.74/0.72 for sparseness, and 0.89/0.68 for firing rate. This shows that one can obtain a good approximation to the optimized populations by selecting neurons based on a biophysical property (encoding time) that may be intrinsically available to cortical circuits.

## Robustness of optimized populations to choice of stimulus ensembles

In the above analyses, populations were optimized for a given stimulus ensemble $S$ consisting of a fixed set of sound epochs. In principle it may be that these populations are highly specific to the respective stimulus ensemble on which they are defined and that they perform poorly when tested on a different ensemble. The above observation that optimized populations generally include the overall informative cells, however, suggests that this should not be the case. Indeed, in an additional control analysis we directly confirmed that the performance of optimized populations tested over stimulus ensembles for which they were not optimized remains close to the performance of the directly optimized population. To this end, we computed 100 optimized populations of size $N = 10$, each optimized on a different stimulus ensemble $S$ ($N_s = 20$). We then tested the performance of each of
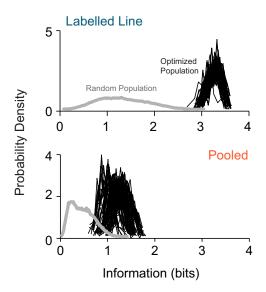
**Figure 5.** Robustness of optimized populations to stimulus selection. Cross-validation of stimulus information for the optimized population on stimulus epochs not used for optimization. One hundred optimized populations of size $N = 10$ were obtained, each from a different stimulus ensemble. For each population, the distribution of information values obtained over the 99 other stimulus ensembles (which were not used in the forward selection procedure) was calculated (black lines) and compared with the distribution of information in 100 random populations over all 100 stimulus ensembles (gray curve). This shows that optimized populations also outperform random populations on stimulus sets not used for their selection (i.e., optimization).

these populations on the other 99 stimulus ensembles that were not used during the forward selection procedure. The results are shown in Figure 5, which illustrates the distributions of information of each optimal population over the other stimulus sets (black lines) together with the distribution of information provided by random populations (gray line). The optimized populations are far outliers from the spread of random populations (minimum *t*-statistics between optimized and random information values are 103.3/27.7 for labeled-line/pooled codes, respectively).

**Trade-offs between population size and temporal precision**
The above shows that optimally selected populations, unlike random populations, provide high amounts of information with few selected neurons. We then investigated whether and how this result depends on the temporal precision ($\Delta t$) used to sample spike trains, varying $\Delta t$ from 10 to 120 ms (Fig. 6). The observation that those units participating in the optimized populations have short encoding time scale already suggests that optimized populations may carry information at high temporal precision (compare Fig. 4). To facilitate comparison across values of $\Delta t$ we used the same optimal population for each stimulus ensemble (derived for $\Delta t = 40$ ms) and tested its decoding performance at other temporal precisions. Figure 6 shows the resulting information values averaged across 100 stimulus ensembles as contours of constant information. This provides several interesting insights.

These results confirm that the optimized populations carry more information and with much fewer neurons and do so for the entire range of analyzed precisions. More importantly, they show that precision and population size trade-off differently in optimized and random populations. For random populations contour lines transverse diagonally, indicating that information lost by decreasing precision can be recovered by adding more

neurons. For the optimized population, however, information lost by decreasing precision cannot be recovered by adding more neurons. For example, the labeled line code can achieve four bits of information at $\Delta t = 10$ ms precision using only $N = 12$ optimally selected units. However, at a temporal precision of $\Delta t = 20$ ms or coarser, such performance levels could not be reached even when considering the full population. Similarly, for the pooled code the maximum obtainable information was limited by the temporal precision, and increasing the number of neurons could not compensate for the loss of temporal precision. To further quantify this trend we performed regression of these information values on precision and population size. This confirmed that precision was relatively more important than population size for the optimized code than the random code ($R^2$ values [population, bin] were [0.29, 0.43] for optimized and [0.87, 0.08] random labeled line codes; and [0.02, 0.89] for optimized and [0.62, 0.31] for random pooled codes). This shows that averaging over random populations can obscure the benefits of exploiting temporally precise responses in a population code.

**Descriptive statistics of population coding**
Several previous studies characterized population codes using specific indices, such as population sparseness and dispersal (Willmore and Tolhurst, 2001; Weliky et al., 2003). For comparison with this important line of work, we computed these quantities for random and optimized populations (for ensemble size of $N_s = 20$ and population size of $N = 20$). Population sparseness characterizes the proportion of neurons being active for any given stimulus. We found that optimized populations were comparably sparse ($0.49 \pm 0.03$; mean $\pm$ SD) to random populations ($0.53 \pm 0.05$) suggesting a similar spread of activity across neurons in each population. Dispersal measures the response variability across the stimulus ensemble for individual neurons and how this is distributed across the population of neurons. Low dispersal arises when few neurons exhibit highly variable responses but most neurons respond similarly to most stimuli, while high dispersal arises when most neurons exhibit a similar variability of responses across stimuli. We found that optimized populations were considerably more dispersed ($0.212 \pm 0.01$) than random populations ($0.145 \pm 0.04$), suggesting that individual neurons contribute more equally in the optimized than within the random populations. Overall, high population sparseness and dispersal are considered key attributes of efficient population codes (Treves and Rolls, 1991; Willmore and Tolhurst, 2001; Weliky et al., 2003), and these numbers suggest that optimized populations better conform to this notion than random populations.

**Discussion**
An established approach to characterizing how the cortical code distributes across cells is to use a combination of experimental data and theoretical models to study how sensory information depends upon population size (Zohary et al., 1994; Stevens and Zador, 1995; Shamir and Sompolinsky, 2006; Cohen and Maunsell, 2009). When doing so it is tempting to consider the average performance of randomly assembled populations to average out potential peculiarities arising from small sets of neurons that may exist within a limited dataset. Such homogenized populations exhibit clear patterns of almost-linear increase of information with population size that can be explained by theoretical models assuming the independent sampling of multiple neurons within the limited entropy of a fixed stimulus set (Abbott et al., 1996; Rolls et al., 1997; Quiroga et al., 2007). This suggests that discrim-
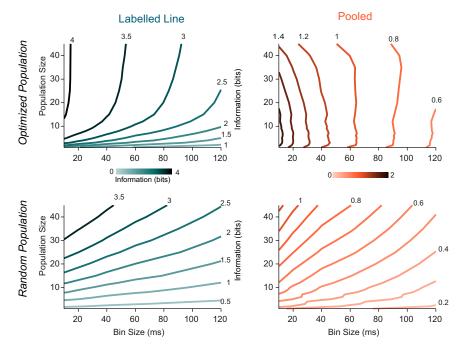
**Figure 6.** Stimulus information as a function of population size and temporal precision. Contour plots of constant information across variations in response precision ($\Delta t$) and population size ($N$). Information values were averaged over 100 stimulus ensembles ($N_s = 20$; $T = 120$ ms). For random populations, the information increases both with increasing population size and decreasing precision. For optimized populations the maximal attainable information is limited by temporal precision.

ination within high dimensional stimulus spaces necessitates monitoring very large populations (Abbott et al., 1996). Our results confirm this theoretical information scaling when considering the average randomly chosen population.

We found, however, that such apparently unbiased averaging may provide a biased view on information coding, as it washes out the privileged contribution of small subsets of neurons. When viewed from the perspective of optimized subpopulations, information saturates much more quickly and essentially all information is available in a small subset of neurons. This implies that the apparent need to monitor large ensembles results from the averaging over random subsets and reflects a biased perspective. We found that the scaling of information in optimized subpopulations can be explained based on the hypothesis that neurons with diverse values of single neuron information randomly cover the stimulus entropy space. This explains the nature of the optimized populations: they are composed of neurons that are highly informative in their own right and that cover different portions of the stimulus space. However, demonstrating the presence of these highly informative subpopulations does not immediately solve the problem of how such populations might be identified and selected for "read-out" by a downstream neural system, but we address this issue below.

The notion that small and privileged populations carry considerable information fits well with the apparent sparseness of cortical activity. Sensory neurons in general feature highly nonuniform and long-tailed distributions of response amplitude across stimuli (Willmore and Tolhurst, 2001; Weliky et al., 2003; Hromádka et al., 2008), a feature known as response sparseness that is considered a sign of computational efficiency (Rolls and Tovee, 1995; Vinje and Gallant, 2000; Olshausen and Field, 2004). This sparseness of individual neuronal responses results in a similarly long-tailed distribution of single neuron selectivity and information (but see Schneidman et al., 2011) and concords

well with the observed overall low population activity in cortical ensembles (Barth and Poulet, 2012). Hence, one interpretation of our findings is that sparse single neuron encoding can create similarly sparse (i.e., small and highly efficient) population codes. While the optimized populations did not exhibit higher population sparseness than random populations, they had higher dispersal, implying a more uniform contribution of the included neurons to the population performance. A remaining challenge is to elucidate the contribution of those neurons that apparently seem silent (Barth and Poulet, 2012) and to fully understand the interplay between firing and silence in shaping population coding (Schneidman et al., 2011).

Experiments have shown that only few cortical sensory neurons are active at any moment in time and that this sparseness is especially prominent in the supragranular layers providing feed-forward connectivity (Greenberg et al., 2008; Hromádka et al., 2008; Histed et al., 2009; Crochet et al., 2011). This suggests that the effective assemblies driving down-stream targets may consist of only few tens of neurons (Tiesinga et al., 2008; Ainsworth et al., 2012; Barth and Poulet, 2012), a hypothesis that is supported by experiments showing that stimulation of small populations can have a sizable impact on network activity and behavior (Brecht et al., 2004; Huber et al., 2008; London et al., 2010). Along this line, a recent study on mouse auditory cortex found that stimulus discrimination can be well explained by the monitoring of a small subset of spatially distributed neurons (Bathellier et al., 2012). In that study highly informative populations could be reduced to low dimensional modes that carried most sensory information. This complementary evidence supports our hypothesis that few units within a large population may be sufficient for sensory encoding.

The use of small populations is computationally attractive. First, implementing a labeled line requires that decoding circuits retain the identity of each afferent. This requires dendritic computations that are biophysically realistic only for limited numbers of afferents (Segev and London, 2000). Second, averaging many afferents is even more destructive for pooled codes. Relying on only a few neurons within a population is hence advantageous regardless of the biophysical mechanisms used to group afferents within a decoding circuit. Third, previous work showed that even weak correlations between neurons can limit the performance of high dimensional population codes (Zohary et al., 1994; Schneidman et al., 2006; Roudi et al., 2009). The use of small populations can avoid the limits imposed by such correlations, which is especially true for heterogonous ensembles of neurons as studied here (Ecker et al., 2011).

Our results also provide insights into the relevance of response timing at the population level. The responses of individual auditory cortical neurons need to be decoded at high temporal precision to recover sensory information and to account for behavioral performance (Schnupp et al., 2006; Engineer et al., 2008; Wang et al., 2008; Kayser et al., 2010; Perez et al., 2013). One may argue that cortical circuits could sacrifice temporal pre-

cision and recover information by monitoring sufficiently many neurons at the same time (Zohary et al., 1994; Shadlen and Newsome, 1998). Our results show that this trade-off is only possible for randomly sampled populations, which, however, provide much less overall information and impose more challenges for read-out than small ensembles. For small and highly informative populations our results show that the temporal precision used by a decoder sets a direct limit on the total recoverable information. This is particularly true for the pooled code, where adding additional neurons may even decrease information due to interference. Our considerations and results, however, do not address per se the question of whether precise spike times of neural populations are used for behavior. This question can be better addressed by comparisons between spike timing information and behavioral performance (Luna et al., 2005; Engineer et al., 2008; Jacobs et al., 2009). However, our results are consistent with a potential role of response timing for population codes, and suggests that an ideal population decoding strategy may be to sample few neurons at the critical resolution for the considered system (Yang and Zador, 2012).

How could a downstream decoder select the optimal neurons to monitor? We found that a simple biophysical marker is sufficient: small optimal populations consist of those neurons with fastest response variations. Such neurons with short encoding time scale often have sparse (Chen et al., 2012) and highly informative responses themselves (Kayser et al., 2010). Selecting neurons with rapid response variations that share little temporal overlap may hence serve as a biophysical marker for informative ensembles that provide independent sensory evidence. Importantly, a selection of neurons based on response dynamics could in principle be implemented by simple synaptic mechanisms. Cortical synapses are equipped with adaptive and plastic mechanisms that are sensitive at different time scales and that can learn to differentiate precise temporal activity patterns even in the presence of background activity (Gütig and Sompolinsky, 2006; Masquelier and Thorpe, 2007).

### Conclusions

Technological advances enable us to simultaneously record from many neurons or even stimulate them, yet understanding the principles of cortical population coding still remains a challenge (Blumhagen et al., 2011; Harvey et al., 2012; Miura et al., 2012). While in some systems the complete recordings of the entire population may be possible (Jacobs et al., 2009), cortical studies still rely on the investigation of subsets of neurons. Future work may benefit from refraining from a random subsampling or extrapolation strategy to create population responses. As we found, this may underestimate the actual coding capacities of a given population or even provide misleading conclusions. Rather, many insights may be obtained by individuating and characterizing the properties of the smallest number of variables describing a population code that yields sufficient information for the task at hand—in analogy to classical definitions of neural codes based on minimal description (Victor, 2000; Panzeri et al., 2010). We found that cortical circuits may recover more sensory information when reading the "right" subset rather than all neurons. Hence, reporting the smallest population that can account for behavior may provide more insights than reporting properties of the average random population. This illustrates the need to carefully consider specific ensembles within a large population under study and to characterize their contributions toward behavior; a challenging task for future work.

## References

Abbott LF, Rolls ET, Tovee MJ (1996) Representational capacity of face coding in monkeys. Cereb Cortex 6:498–505. CrossRef Medline

Ainsworth M, Lee S, Cunningham MO, Traub RD, Kopell NJ, Whittington MA (2012) Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks. Neuron 75:572–583. CrossRef Medline

Alpaydin E (2010) Introduction to machine learning, Ed 2. Cambridge, MA: MIT.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366. CrossRef Medline

Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, Wakeman EA, Rolls ET (1997) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proc Biol Sci 264:1775–1783. CrossRef Medline

Barth AL, Poulet JF (2012) Experimental evidence for sparse firing in the neocortex. Trends Neurosci 35:345–355. CrossRef Medline

Bathellier B, Ushakova L, Rumpel S (2012) Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron 76:435–449. CrossRef Medline

Blumhagen F, Zhu P, Shum J, Schärer YP, Yaksi E, Deisseroth K, Friedrich RW (2011) Neuronal filtering of multiplexed odour representations. Nature 479:493–498. CrossRef Medline

Brecht M, Schneider M, Sakmann B, Margrie TW (2004) Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. Nature 427:704–710. CrossRef Medline

Chen C, Read HL, Escabí MA (2012) Precise feature based time scales and frequency decorrelation lead to a sparse auditory code. J Neurosci 32:8454–8468. CrossRef Medline

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594–1600. CrossRef Medline

Crochet S, Poulet JF, Kremer Y, Petersen CC (2011) Synaptic mechanisms underlying sparse coding of active touch. Neuron 69:1160–1175. CrossRef Medline

Douglas RJ, Martin KA (2007) Mapping the matrix: the ways of neocortex. Neuron 56:226–238. CrossRef Medline

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. Science 327:584–587. CrossRef Medline

Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. J Neurosci 31:14272–14283. CrossRef Medline

Engineer CT, Perez CA, Chen YH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP (2008) Cortical activity patterns predict speech discrimination ability. Nat Neurosci 11:603–608. CrossRef Medline

Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? J Neurosci 13:2758–2771. Medline

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. J Neurophysiol 76:1356–1360. Medline

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535–574. CrossRef Medline

Graf AB, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. Nat Neurosci 14:239–245. CrossRef Medline

Greenberg DS, Houweling AR, Kerr JN (2008) Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat Neurosci 11:749–751. CrossRef Medline

Gütig R, Sompolinsky H (2006) The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci 9:420–428. CrossRef Medline

Harvey CD, Coen P, Tank DW (2012) Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484:62–68. CrossRef Medline

Hastie T, Tibshirani RJ, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. New York: Springer.

Histed MH, Bonin V, Reid RC (2009) Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. Neuron 63:508–522. CrossRef Medline

Hromádka T, Deweese MR, Zador AM (2008) Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol 6:e16. CrossRef Medline

Huber D, Petreanu L, Ghitani N, Ranade S, Hromádka T, Mainen Z, Svoboda K (2008) Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. Nature 451:61–64. CrossRef Medline

Jacobs AL, Fridman G, Douglas RM, Alam NM, Latham PE, Prusky GT, Nirenberg S (2009) Ruling out and ruling in neural codes. Proc Natl Acad Sci U S A 106:5936–5941. CrossRef Medline

Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61:597–608. CrossRef Medline

Kayser C, Logothetis NK, Panzeri S (2010) The millisecond encoding precision of auditory cortex neurons. Proc Natl Acad Sci U S A 107:16976–16981. CrossRef Medline

Kayser C, Ince RA, Panzeri S (2012) Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. PLoS Comput Biol 8:e1002717. CrossRef Medline

Kwan AC, Dan Y (2012) Dissection of cortical microcircuits by single-neuron stimulation *in vivo*. Curr Biol 22:1459–1467. CrossRef Medline

Lehky SR, Kiani R, Esteky H, Tanaka K (2011) Statistics of visual responses in primate inferotemporal cortex to object stimuli. J Neurophysiol 106:1097–1117. CrossRef Medline

Logothetis NK, Augath M, Murayama Y, Rauch A, Sultan F, Goense J, Oeltermann A, Merkle H (2010) The effects of electrical microstimulation on cortical signal propagation. Nat Neurosci 13:1283–1291. CrossRef Medline

London M, Roth A, Beeren L, Häusser M, Latham PE (2010) Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. Nature 466:123–127. CrossRef Medline

Lu T, Wang X (2004) Information content of auditory cortical responses to time-varying acoustic stimuli. J Neurophysiol 91:301–313. Medline

Luna R, Hernández A, Brody CD, Romo R (2005) Neural codes for perceptual discrimination in primary somatosensory cortex. Nat Neurosci 8:1210–1219. CrossRef Medline

Magri C, Whittingstall K, Singh V, Logothetis NK, Panzeri S (2009) A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. BMC Neurosci 10:81. CrossRef Medline

Masquelier T, Thorpe SJ (2007) Unsupervised learning of visual features through spike timing dependent plasticity. PLoS Comput Biol 3:e31. CrossRef Medline

Miura K, Mainen ZF, Uchida N (2012) Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. Neuron 74:1087–1098. CrossRef Medline

Narayanan NS, Kimchi EY, Laubach M (2005) Redundancy and synergy of neuronal ensembles in motor cortex. J Neurosci 25:4207–4216. CrossRef Medline

Nelken I, Chechik G (2007) Information theory in auditory research. Hear Res 229:94–105. CrossRef Medline

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487. CrossRef Medline

Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. Network 7:87–107. CrossRef

Panzeri S, Schultz SR, Treves A, Rolls ET (1999) Correlations and the encoding of information in the nervous system. Proc Biol Sci 266:1001–1012. CrossRef Medline

Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33:111–120. CrossRef Medline

Perez CA, Engineer CT, Jakkamsetti V, Carraway RS, Perry MS, Kilgard MP (2013) Different timescales for the neural coding of consonant and vowel sounds. Cereb Cortex 23:670–683. CrossRef Medline

Perkel DH, Bullock TH (1968) Neural coding. Neurosci Res Progr Bull 6:221–348.

Pouget A, Dayan P, Zemel RS (2003) Inference and computation with population codes. Annu Rev Neurosci 26:381–410. CrossRef Medline

Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173–185. CrossRef Medline

Quiroga RQ, Reddy L, Koch C, Fried I (2007) Decoding visual inputs from multiple neurons in the human temporal lobe. J Neurophysiol 98:1997–2007. CrossRef Medline

Reich DS, Mechler F, Victor JD (2001) Independent and redundant information in nearby cortical neurons. Science 294:2566–2568. CrossRef Medline

Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726. Medline

Rolls ET, Treves A, Tovee MJ (1997) The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. Exp Brain Res 114:149–162. CrossRef Medline

Roudi Y, Nirenberg S, Latham PE (2009) Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. PLoS Comput Biol 5:e1000380. CrossRef Medline

Safaai H, von Heimendahl M, Sorando JM, Diamond ME, Maravall M (2013) Coordinated population activity underlying texture discrimination in rat barrel cortex. J Neurosci 33:5843–5855. CrossRef Medline

Schneidman E, Bialek W, Berry MJ 2nd (2003) Synergy, redundancy, and independence in population codes. J Neurosci 23:11539–11553. Medline

Schneidman E, Berry MJ 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007–1012. CrossRef Medline

Schneidman E, Puchalla JL, Segev R, Harris RA, Bialek W, Berry MJ 2nd (2011) Synergy from silence in a combinatorial neural code. J Neurosci 31:15732–15741. CrossRef Medline

Schnupp JW, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26:4785–4795. CrossRef Medline

Segev I, London M (2000) Untangling dendrites with quantitative models. Science 290:744–750. CrossRef Medline

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870–3896. Medline

Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. Neural Comput 18:1951–1986. CrossRef Medline

Stevens CF, Zador AM (1995) Neural coding: the engima of the brain. Curr Biol 12:1370–1371.

Theunissen F, Miller JP (1995) Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci 2:149–162. CrossRef Medline

Tiesinga P, Fellous JM, Sejnowski TJ (2008) Regulation of spike timing in visual cortical circuits. Nat Rev Neurosci 9:97–107. CrossRef Medline

Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? Network 2:371–397. CrossRef

Victor JD (2000) How the brain uses time to represent and process visual information. Brain Res 886:33–46. CrossRef Medline

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276. CrossRef Medline

Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 157:484–494. CrossRef Medline

Weatherall D (2006) The Weatherall report on the use of non-human primates in research. London: The Royal Society.

Weliky M, Fiser J, Hunt RH, Wagner DN (2003) Coding of natural scenes in primary visual cortex. Neuron 37:703–718. CrossRef Medline

Willmore B, Tolhurst DJ (2001) Characterizing the sparseness of neural codes. Network 12:255–270. CrossRef Medline

Willmore BD, Mazer JA, Gallant JL (2011) Sparse coding in striate and extrastriate visual cortex. J Neurophysiol 105:2907–2919. CrossRef Medline

Yang Y, Zador AM (2012) Differences in sensitivity to neural timing among cortical areas. J Neurosci 32:15142–15147. CrossRef Medline

Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143. CrossRef Medline