Supporting Information

Clemens et al. 10.1073/pnas.1104506108

SI Results

Specificity of Spike-Triggered Averages. In Fig. 3 in the main text, we showed that spike-triggered averages (STAs)—and hence the feature selectivity—become more diverse at the level of ascending neurons. The decoding approach (Fig. 4 in the main text) led us to conclude that ascending neurons are optimally read out as a labeled line, indicating that each ascending neuron signals a specific aspect of the stimulus.

The second and third processing layers each consist of several cell types. To further support our findings, we also looked at the similarity of each cell type's STA across different individuals/recordings by computing Pearson's correlation coefficient between the average STA of a cell type and the STA of each individual cell of that type. In order for the STA (feature) to be specific for the cell type, this "intratype" similarity should be larger than the similarity across different cell types of that layer ("intertype" similarity).

As we considered each receptor a different type, inter- and intratype similarities for receptors are identical (Fig. S3, blue box plots). For local neurons (Fig. S3, red box plots), both the STAs of the same cell type and of different cell types are highly similar (intratype similarity 0.95 ± 0.04, mean $\pm$ SD; intertype similarity 0.81 ± 0.10). Whereas individual STAs of the same type are significantly more similar than those of different types ($P = 0.019$, rank-sum test), overall similarity at the level of local neurons is high. In ascending neurons, however, STAs of the same cell type are much more similar than those of different types (intratype similarity 0.85 ± 0.17, intertype similarity 0.22 ± 0.55; $P = 0.002$, rank-sum test). This cell specificity of STA filters further supports our hypothesis that each type of ascending neuron encodes a specific aspect of the stimulus.

Information Gain Relative to the Population Average. In the main text, we quantified the information gain by relating the information of a four-cell population (the larger value of those obtained with the summed-population and the labeled-line decoder) with that of the best cell in that population. Alternatively, we also considered the gain with respect to the average information of all four cells comprising that population (Fig. S5*A*). Clearly, this measure of information gain yields higher values: Receptors exhibit an average gain of 1.99, local neurons 1.49, and ascending neurons 2.49. Thus, the gain relative to the average information in the population is 1.4- to 1.7-fold greater than the gain relative to the best cell. This is due to an upward bias in this alternative measure: The more cells one includes in a population, the more likely it is to "hit" a highly informative one. The receptors with their high spread of single-neuron information values (Fig. 2*D* in the main text) are especially susceptible to this bias. We hence decided to quantify information gain relative to the best cell in each population as a more conservative and less biased measure.

SI Materials and Methods

Decoding. We quantified information in neural responses using a decoding approach (1). Although we thereby underestimate the full information in the statistical sense, we probably come closer to what a concrete, biologically plausible system can read out from the spike trains we study here.

Single-neuron metric. The spike-train dissimilarity of single neurons was quantified using the van Rossum metric (2). Spike trains were binned with a resolution of 0.05 ms and filtered with an α function: $\alpha(t) = \Theta(t)\, t \exp(-t/\tau)$, where $\Theta(t)$ is Heaviside's function. The parameter τ governs the temporal resolution of the metric. The Euclidean distance between all pairs of responses (eight repetitions of eight song segments of different males, duration 200 ms each) yields a distance matrix that forms the base for the classification algorithm outlined below.

Multineuron metric. Population data were combined from single-cell recordings of four individual cells. This was justified, as neural activity in the network is entirely stimulus-driven. Hence, neurons are conditionally independent: There are no "noise" correlations between neurons, only signal correlations (3). Because we were interested in how the population code changed between processing stages, we created three different classes of four-cell populations, combining different types of either receptors or local or ascending neurons. Thus, each population was characterized by a unique combination of four different cell types of a single layer. So as not to overrepresent those populations that consist of cell types we have recorded more often, we averaged information rates and gains for each kind of population (i.e., combination of cell types) for plotting and statistics.

For a formal derivation of the multineuron metric, see ref. 4. Application of this metric amounts to filtering the spike trains with an α function, embedding the spike trains from multiple cells into a vector space, and then taking the Euclidean distance between different spike trains. The resulting distance matrix for each population is then used to quantify stimulus discriminability through the classification algorithm. Thus, the only difference from the single-cell metric is that the spike trains of the cells comprising a population are embedded in a vector space.

The multineuron metric allows for different kinds of embedding, which is controlled by the "independence" parameter θ—the "angle between cells." This parameter allows interpolating between two versions of a population code: a summed-population code and a labeled-line code. At $\theta = 0°$, the metric corresponds to a summed-population code, where responses of different cells are embedded colinearly. Information about which cell fired which spike is lost. This is optimal only if differences in the firing pattern between cells in a population are not relevant for the decoding tasks or if cells in a population are similarly tuned—this applies in our case to receptors and local neurons. In contrast, information about each spike's origin is fully retained in a labeled-line code, which is implemented at $\theta = 90°$ (orthogonal embedding). This is desirable, if cells are tuned differently and represent different aspects of a stimulus, like the ascending neurons.

To illustrate that the labeled-line decoder incorporates information about which neuron fired which spikes—the neuronal identity of spikes—we provide a simplified example of how three different stimuli can be distinguished with the summed-population and the label-line decoder, respectively, based on surrogate responses from two neurons. Fig. S4*A* shows the surrogate spike trains of both cells in response to the three different stimuli. To simplify the argument and without loss of generality, we reduce these spike trains to spike counts, which corresponds to applying a filter with a large time constant τ. In response to stimulus 1, cell A (green) and cell B (blue) fire three spikes each. Stimulus 2 evokes one spike in cell A and five spikes in cell B. The response pattern for stimulus 3 is inverted: Now cell A fires five spikes and cell B only one.

The summed-population decoder sums these spike counts before computing pairwise distances between all stimuli. As the sum of spikes in cell A and cell B is the same, the population response to all three stimuli is represented by a 6; they cannot be distinguished. In contrast, the labeled-line decoder does not pool

the response of the two cells in the population. Here each response is represented by an ordered pair of spike counts, which is different for each stimulus. This is also reflected in the resulting distance matrices (Fig. S4*B*). As the summed-population spike counts are the same for all three stimuli, the distance matrix has all zero entries and the summed-population decoder cannot discriminate between the three stimuli (information 0 bit). The labeled-line decoder, however, discriminates all three stimuli, as all off-diagonal entries in the distance matrix exhibit nonzero entries (information $log_2 3 = 1.6$ bit). The labeled-line decoder can distinguish stimulus 1 from both stimuli 2 and 3. In particular, it can also disambiguate stimuli 2 and 3, which differ only in neuronal identity of responses (both stimuli evoke one spike in one cell and five in the other, but in a different order). This ordering is the major difference for the summed-population decoder, and reflects the role of neuronal identity for the labeled-line decoder.

Classifier. Responses were classified using a nearest-neighbor clustering algorithm as in ref. 5. Nearness was given by the single or multineuron metrics. We randomly selected one template spike train from each of the eight songs. The remaining spike trains were then classified as being evoked by the song to which the nearest template belonged. This was repeated many times, always with a new, randomly selected set of templates. We organized the classification results in a confusion matrix $H(s,s')$, which shows the frequency with which a spike train being evoked by song s was classified as being evoked by song s'. The average of this matrix's main diagonal denotes the fraction of correctly decoded spike trains.

Estimation of information. The mutual information of this confusion matrix $I(s,s')$ was used as a proxy for the information content of the neural responses $I(s,r)$ (1). Information is given by $I(s,r) \propto I(s,s') = \sum_{s,s'} p(s,s') \log_2 \frac{p(s,s')}{p(s)p(s')}$, where $p(s,s')$ is the entry in the confusion matrix, $p(s) = \sum_{s'} p(s,s') = 1/8$ is the prior stimulus probability, and $p(s') = \sum_{s} p(s,s')$ is the marginal over the decoded stimuli (6). Mutual information is 0 bit when the confusion matrix is uniformly distributed, that is, when each entry has the value 1/64. It is maximal [for eight stimuli $log_2(8) = 3$ bit] when there is a one-to-one relationship between spike trains and classes, for example, when all entries are concentrated at the matrix's diagonal. As this measure is upwardly biased, we calculated the same quantity 10 times for random assignments between responses and stimulus classes and subtracted this bias from the naive estimator $I(s,s')$ (7). We expressed information either as a rate in bit/s by dividing the information by the stimulus length (maximal information rate being thereby 8/0.2 s = 15 bit/s) or as information per spike (bit per spike) by normalizing the information rate by the cell's firing rate. Firing rate was quantified as the spike count divided by the length of the spike train segment (200 ms).

Optimization of the metric's parameters. Classification performance is a function of the metric's temporal resolution τ. We optimized information with a grid search for τ ranging from 0.25 to 64 ms (nine values, spaced linearly on a logarithmic scale). The τ used for decoding are shown in Fig. S6. Receptors exhibited an intermediate range of τ between 4 and 8 ms with two outliers at 16 and 32 ms. The τ of local neurons were significantly smaller ($P = 0.01$), spanning a range of 3–4.2 ms. Ascending neurons had the highest τ between 6.7 and 42 ms, being significantly greater than those of local neurons ($P = 0.003$). For population decoding with the multineuron metric, we used a single optimal τ for all cells in a population.

In the main text, we consider only the information rates obtained for two "extreme-value decoders" at $\theta = 0°$ (summed-population) and at $\theta = 90°$ (labeled-line) for each population. We have also determined information at the *optimal* θ for each population by a grid search in the interval [0°, 90°]. As either of the two decoders at 0° or 90° yielded near-optimal performance for any population (median information loss 2%), we decided to consider only those two for all analyses.

Statistics. All plots and statistics were based on average values for each cell type or type of population, that is, over all recordings of a cell type for the analysis of single cells and over unique, unordered 4-tuples for populations of cells. Tests—if not stated otherwise—were either parametric (*t* test) or nonparametric (two-sided Wilcoxon's rank-sum test), depending on the outcome of a Jarque–Bera test for normality with a significance level $\alpha = 0.05$. No correction for multiple comparisons was performed to avoid false negatives, as we were interested in the outcome of each individual pairwise comparison, not in the general detection of statistical differences between groups.

All analysis was done in MATLAB (The MathWorks).

1. Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: Information theory and decoding approaches. *Nat Rev Neurosci* 10:173–185.
2. van Rossum MC (2001) A novel spike distance. *Neural Comput* 13:751–763.
3. Brody CD (1999) Disambiguating different covariation types. *Neural Comput* 11: 1527–1535.
4. Houghton CJ, Sen K (2008) A new multineuron spike train metric. *Neural Comput* 20: 1495–1511.
5. Machens CK, et al. (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci* 6:341–342.
6. Victor JD, Purpura K (1997) Metric-space analysis of spike trains: Theory, algorithms and application. *Network* 8:127–164.
7. Aronov D, Reich DS, Mechler F, Victor JD (2003) Neural coding of spatial phase in V1 of the macaque monkey. *J Neurophysiol* 89:3304–3327.

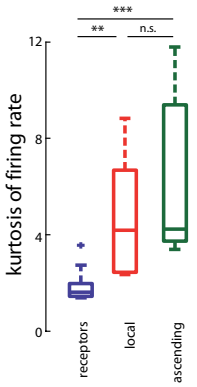

Fig. S1. Kurtosis of the firing-rate distribution as an alternative measure of lifetime sparseness. Receptors 1.9 ± 0.7, local neurons 4.8 ± 2.7, ascending neurons 6.2 ± 3.6; receptors versus local neurons $P = 0.008$, receptors versus ascending neurons $P = 2.1 \times 10^{-4}$, local neurons versus ascending neurons $P = 0.5$. n.s., nonsignificant. $P > 0.05$, $^{**}P > 0.01$, $^{***}P < 0.001$, rank-sum test.

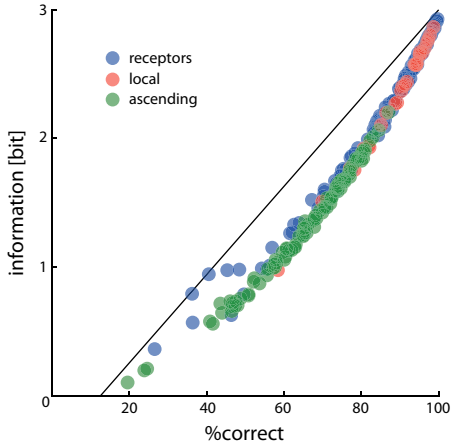

Fig. S2. Two measures of decoder performance are highly correlated. In our dataset, both the mutual information and the percentage of correct classification yield highly similar results for decoding of single cells as well as of populations ($r^2 = 0.97$).

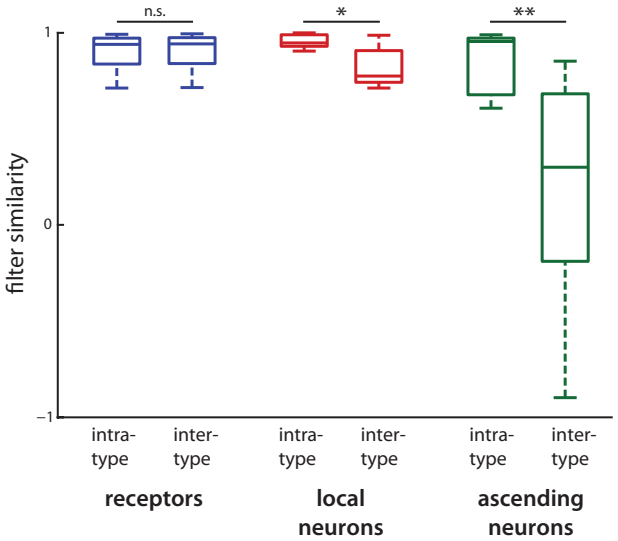

Fig. S3. Cell-type specificity of STA filters. Shown is the similarity of the STA filters of different specimens of the same cell type (intratype) and the similarity of the STA filters of different cell types (intertype; same as Fig. 3*C* in the main text). n.s., nonsignificant. $P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$, rank-sum test.

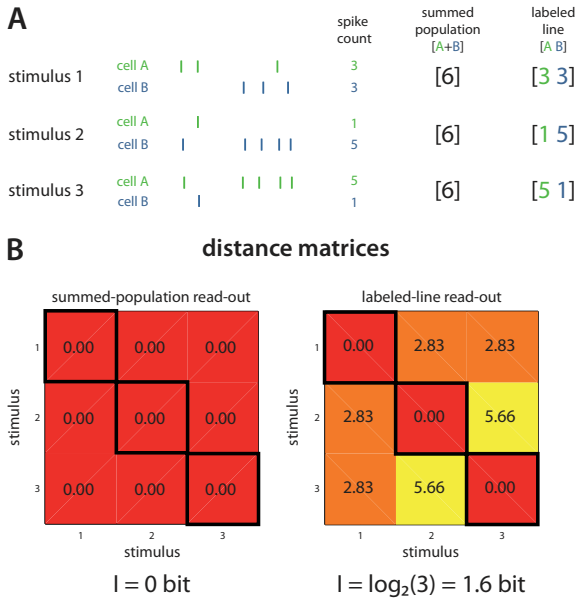

Fig. S4. Illustration of the summed-population and the labeled-line decoder. (*A*) Artificially generated responses of two cells (A and B) to three arbitrary stimuli (*Left*). (*Right*) A representation of the spike counts by the summed-population and labeled-line decoder is shown. (*B*) Resulting distance matrices and information values.

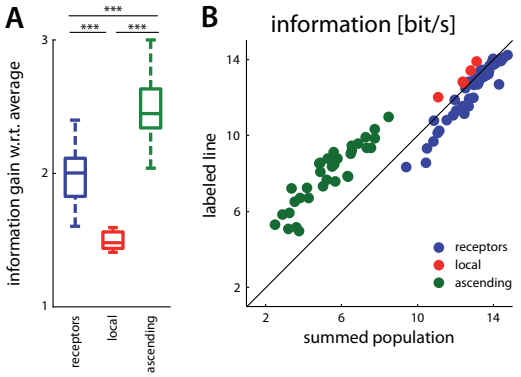

Fig. S5. Information of the population decoder. (*A*) Information gain of the population with respect to the average information obtained by decoding them individually. Receptors 2.0 ± 0.2, local neurons 1.5 ± 0.1, ascending neurons 2.5 ± 0.2; all $P < 6 \times 10^{-7}$, rank-sum test. Horizontal line at 1.0 indicates no gain. ***$P < 0.001$, rank-sum test. (*B*) Information of the summed-population versus that of the labeled-line decoder. Results for decoder performance in terms of % correct look similar (*Inset* in Fig. 4*B* in the main text). Blue, receptors (n = 100 four-cell populations); red, local neurons (n = 160 four-cell populations, 5 different combinations of cell types); green, ascending neurons (n = 910 four-cell populations, 35 different combinations of cell types).

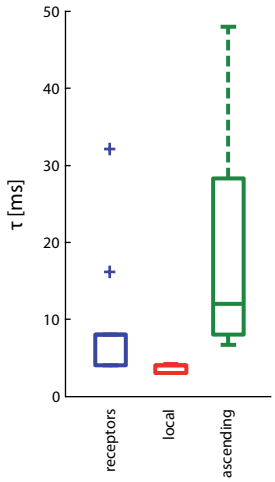

Fig. S6. Optimal timescales for decoding. Box plots show the τs that maximized the mutual information for each cell type. These determine the width of the α functions with which spike trains were convolved in the decoding procedure and indicate the timescale at which the decoder operated.