Neuron
Article

Discrete Neocortical Dynamics Predict Behavioral Categorization of Sounds

Brice Bathellier,[1] Lyubov Ushakova,[1] and Simon Rumpel[1,*]
[1]Research Institute of Molecular Pathology (IMP), Dr Bohr-Gasse 7, 1030 Vienna, Austria
*Correspondence: rumpel@imp.ac.at
http://dx.doi.org/10.1016/j.neuron.2012.07.008

SUMMARY

The ability to group stimuli into perceptual categories is essential for efficient interaction with the environment. Discrete dynamics that emerge in brain networks are believed to be the neuronal correlate of category formation. Observations of such dynamics have recently been made; however, it is still unresolved if they actually match perceptual categories. Using in vivo two-photon calcium imaging in the auditory cortex of mice, we show that local network activity evoked by sounds is constrained to few response modes. Transitions between response modes are characterized by an abrupt switch, indicating attractor-like, discrete dynamics. Moreover, we show that local cortical responses quantitatively predict discrimination performance and spontaneous categorization of sounds in behaving mice. Our results therefore demonstrate that local nonlinear dynamics in the auditory cortex generate spontaneous sound categories which can be selected for behavioral or perceptual decisions.

INTRODUCTION

Much of the work on the auditory cortex (AC) has been focused on the analysis of single neuron receptive fields—testing the idea that cortical neurons function as an array of linear filters that decompose sounds in a similar way to the spectrograms used for graphical sound representation. However, recent studies have accumulated evidence that single neurons do not behave as true linear filters (Christianson et al., 2008; David et al., 2009; Machens et al., 2004). Specifically, measures of the linear response characteristics of single neurons to sound (e.g., tuning curve, spectrotemporal receptive field) show that neuronal responses depend on the intensity, the sequence (Christianson et al., 2011; Ulanovsky et al., 2004), and the context of the tested sound (Eggermont, 2011; Nelken et al., 1999; Rabinowitz et al., 2011) as well as on the state of the animal (Atiani et al., 2009). Starting from the theoretical work of J. Hopfield on attractors in recurrent neuronal networks (Hopfield, 1982), modeling studies suggested that cortical-like network architectures are prone to generate highly nonlinear population dynamics (Amit and Brunel, 1997; Maass et al., 2007; Mongillo et al., 2008; Wang, 2008). This highly nonlinear population dynamic could explain the shortcomings of the linear filter model as recently suggested in a model of the AC (Loebel et al., 2007). Importantly, the all-or-none properties of nonlinear population dynamics could serve as a basis for encoding the perceptual categories, or objects, which are essential for efficient and robust interaction with the environment (Miller et al., 2003; Russ et al., 2007; Seger and Miller, 2010). This idea is supported by recent experiments in the rat hippocampus and the zebrafish olfactory bulb reporting abrupt transitions in the neuronal representation of continuously changing olfactory stimuli or spatial environments (Niessing and Friedrich, 2010; Wills et al., 2005). Nonetheless, it remains unclear how far these discrete network dynamics actually reflect perceptual categories since the experimental designs did not involve any perceptual judgment of the stimuli. Here, we investigated whether nonlinear representations of sounds exist in the mouse auditory cortex at the network level and whether they could form a basis for sound categorization, which has been proposed to be a major function of the AC (Nelken et al., 2003; Ohl et al., 2001; Russ et al., 2007).

RESULTS

Temporally Deconvolved Calcium Imaging in the Mouse Auditory Cortex

To monitor network activity in the auditory cortex of the mouse with single cell resolution, we used two-photon calcium imaging, a technique which gives the possibility to simultaneously record the activity of a large number of neurons in vivo (Garaschuk et al., 2006). We injected isoflurane anaesthetized mice (1%) with the calcium-sensitive dye Oregon Green Bapta 1 AM (OGB1) in the region functionally identified as the AC using intrinsic imaging recordings (Figures 1A and 1B; Kalatsky et al., 2005). Neurons labeled with OGB1 were imaged using two-photon microscopy in single focal planes at a depth of ~150–300 μm below the pia in cortical layers II/III (Figure 1C). The typical field of view was a 200 μm square, in which calcium signals from 46–99 individual neurons were recorded using line scans (Figures 1C and 1D). To estimate the neuronal firing rate based on OGB1 fluorescence measurements, we performed loose-patch recordings of individual OGB1 loaded neurons in vivo. The electrically recorded neuron was simultaneously imaged together with its neighbors using our typical line scan settings (Figures 1E and 1F). Consistently with a previous report (Yaksi and Friedrich, 2006), we observe that the temporal deconvolution of the raw calcium signals using an exponential kernel matched the time course of the neuron's instantaneous firing rate (Figures 1G–1H, and see Figure S1 available online). An estimate of the absolute firing

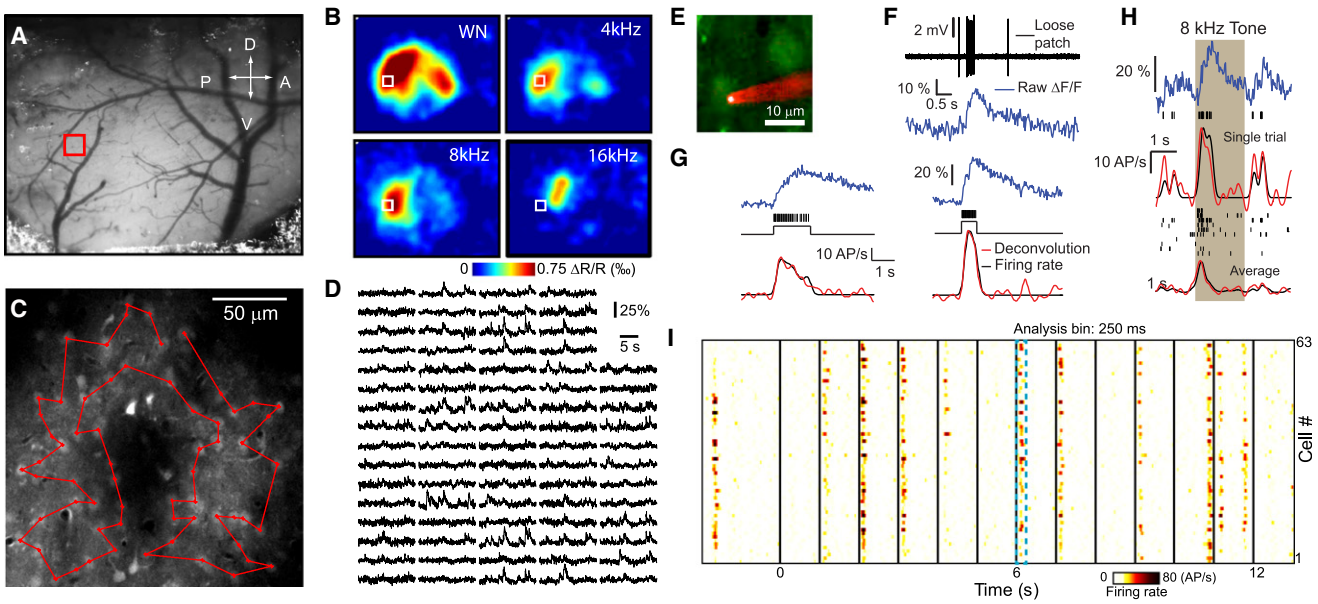

Figure 1. Temporally Deconvolved Calcium Imaging in the Mouse Auditory Cortex

(A) Image of the brain surface with vessel pattern. Red box corresponds to white boxes in (B) and indicates area shown in (C) imaged using two-photon microscopy.

(B) Intrinsic imaging responses to white noise and 3 pure tones. The white squares indicate the location and extent of a calcium recording. A: anterior; P: posterior; V: ventral; D: dorsal.

(C) Image of OGB1 stained neurons at the location shown in (A). The line scan path is superimposed.

(D) Typical calcium signals from all 63 neurons shown in (C).

(E) Two-photon microscopy image of layer 2/3 neurons stained with OGB1 (green). One neuron is loosed-patched with an electrode filled with Sulforhodamine (red).

(F) Example of simultaneous extracellular recording (black trace) and calcium imaging (blue trace) for the cell shown in (E).

(G) Accuracy of temporal deconvolution in the neuron shown in (E). The recorded neuron was extracellularly driven by large current ejections. (Blue lines) Raw calcium signal. (Vertical bars) Time points of electrically recorded spikes. (Gray line) Time course of ejected currents. (Black line) Instantaneous firing rate of the neuron filtered by a Gaussian kernel (σ = 100 ms). (Red line) Deconvolved calcium signal filtered by the same Gaussian kernel.

(H) (Top) Example of a different neuron simultaneously imaged and loose-cell patched during a single presentation of an 8 kHz tone. (Bottom) Raster plot for 10 repetitions of the 8 kHz tone for the same neuron. The red and black curves corresponds to the averaged firing rate and deconvolved calcium signal filtered by a Gaussian kernel (σ = 100 ms).

(I) Firing rate estimate for each recorded neuron derived from the traces shown in (D) after temporal deconvolution. Sound stimuli presented at 1 s intervals are indicated by black vertical bars.

See also Figure S1.

rate amplitude was obtained by linearly scaling the deconvolved signal to fit the actual firing rate. The average scaling factor corresponding to the change in fluorescence elicited by a single action potential across all recordings was 1.80% ± 0.44% (mean ± SD, n = 5).

Sounds Evoke Stochastic Local Responses in the Auditory Cortex

Typical spontaneous and sound evoked AC activity was dominated by short population events in which a large fraction of neurons fired synchronously (Figure 1I). This observation is in agreement with previous reports based on multisite or intracellular current recordings (DeWeese and Zador, 2006; Luczak et al., 2009; Sakata and Harris, 2009). Additionally, it is consistent with the high noise correlations between neurons observed in previous calcium imaging studies (Bandyopadhyay et al., 2010; Rothschild et al., 2010). To evaluate qualitatively how different sounds might generate different types of local population events, we plotted single trial response vectors (~15–20 trials per sound) obtained by averaging the activity for each neuron in a 250 ms time window following sound onset. An example of such plots for four distinct short pure tones (50 ms) at different sound levels is shown in Figure 2A. We observed that all studied populations had local response vectors that were highly variable from one trial to the other for a given sound. This variability was due to large trial-to-trial variations in the response of most individual neurons (Bartho et al., 2009; Hromádka et al., 2008). Despite the variability, we were able to observe that sound intensity and identity modulated the probability of observing a population event. Tuning to pure tones could be seen at both the single neuron (Figures 2B–2D) and at the population level (Figure 2A). However, prediction of the population firing rate in response to complex sounds by a linear model based on the observed pure tone tuning was poor (Figure S2). Therefore, local populations encode sounds in a nonlinear fashion, as was reported for single neurons (Machens et al., 2004). This implies that pure tone tuning alone is not sufficient to describe sound representation in the auditory

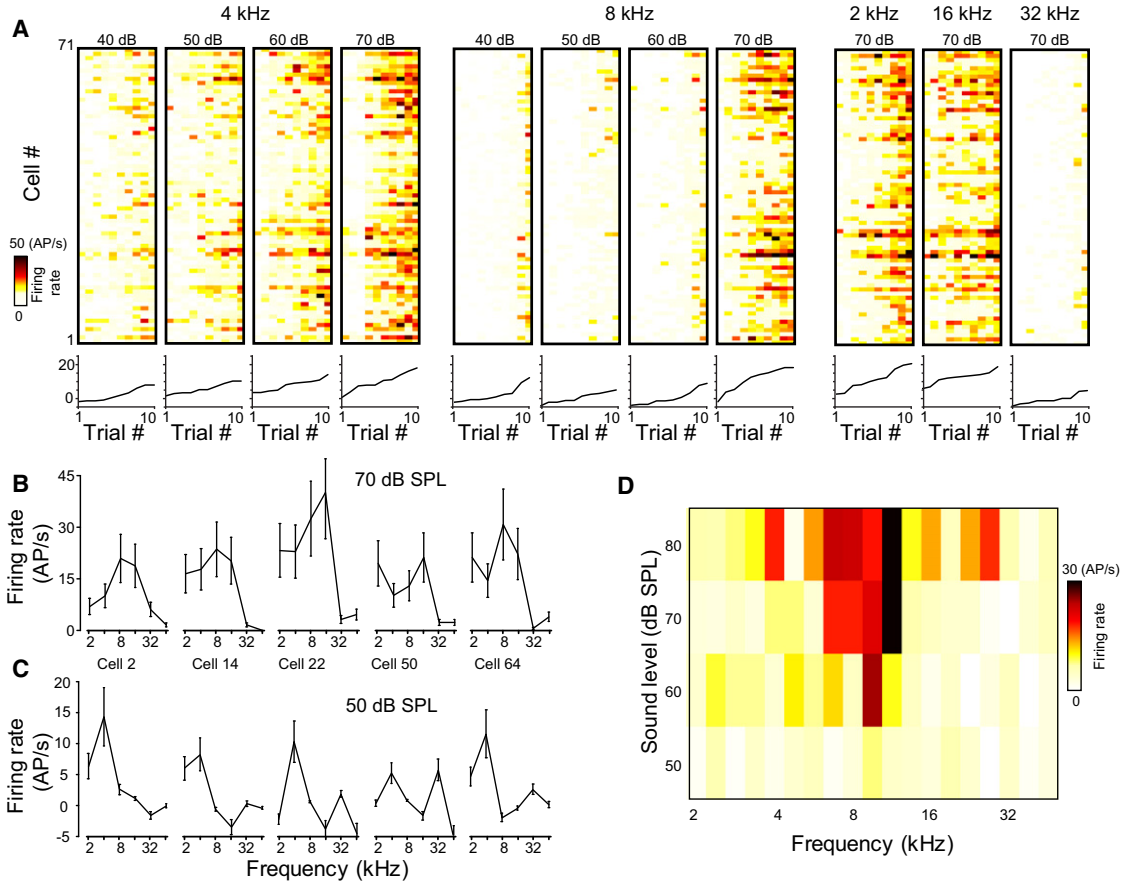

Figure 2. Tuning of Local Population Responses to Pure Tones

(A) Color plot of 10 single trial population responses to pure tone pips of various frequencies indicated above graph. Responses are sorted by increasing mean population firing rate as plotted below. The color code indicates the mean firing rate of a neuron during a 250 ms time bin that starts with sound onset. In this population similar response patterns are evoked by different pure tones of various frequencies and intensities, i.e., the activity profile across the population is largely preserved in trials when a population response is observed.

(B and C) Examples of single cell tuning curves obtained at 2 different sound intensities for 5 different cells of the population shown in (A). Note that the vertical scale is different in (B) and (C). Error bars: SEM.

(D) Example of a frequency response area plot of a single neuron taken from a different recording and showing the typical V-shaped frequency response areas of AC neurons.

See also Figure S2.

cortex. We therefore decided to use a more general framework to investigate the coding properties of local response patterns in single trials (Bathellier et al., 2008). In most local populations, we made the striking observation that despite the high variability of response patterns, the most reliable part of the pattern seemed to be common to very different sounds such as the different pure tones shown in Figure 2A. This suggested to us that sound evoked responses in local auditory cortex networks are constrained to a limited repertoire of functional patterns superposed on high trial-to-trial stochasticity.

Local Network Responses Are Constrained to Few Response Modes

To obtain a quantitative account of the limited repertoire of functional patterns in the face of large variability, we systematically quantified the similarity of local response patterns elicited by large arrays of short (50–70 ms) pure tones and complex sounds. To do so, we used a similarity metric designed to obtain an intuitive readout of single trial response separability. In short, the similarity between two sound-evoked responses was defined as the average of all pairwise correlations between the single trial response patterns of the *two* sounds (see matrices of single trial correlations, Figure 3A). This similarity metric was compared to a response reliability metric, which was the average of all pairwise correlations between all the single trial response patterns of *one* given sound. This reliability metric gave us a quantitative readout of the trial-to-trial variability in response to a given sound. Using these two metrics, the idea is that if the response patterns to two sounds have a lower similarity than their respective reliabilities, they will likely be discriminable on a single trial basis by an external observer. If not, the patterns can be thought to be the same. Pairwise response similarities

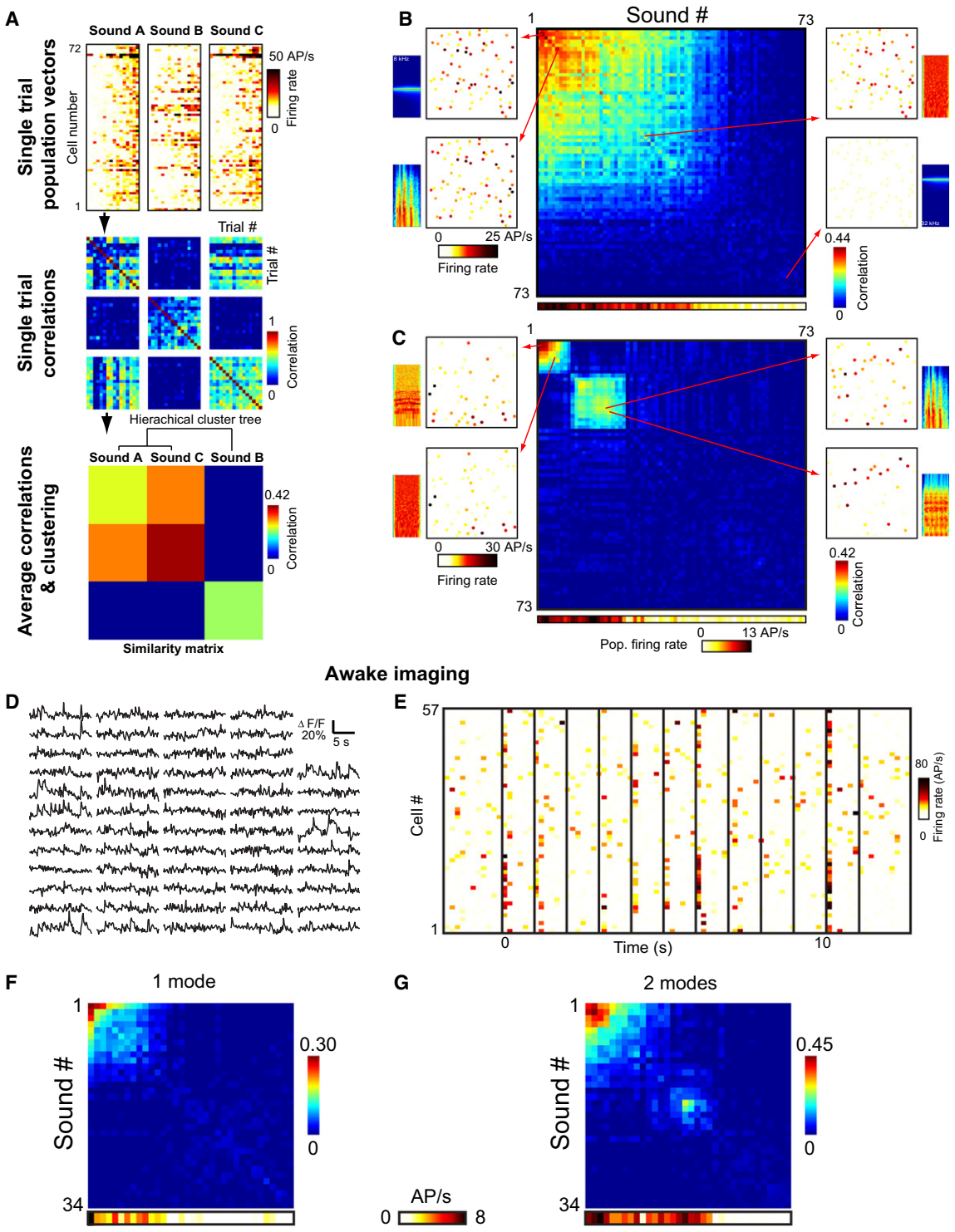

Figure 3. Clustering of Local Population Responses Reveals a Low Number of Response Modes

(A) Construction of a similarity matrix based on the average of the correlation coefficients of all single trial combinations followed by hierarchical clustering from population responses in a local population evoked by three different sounds.

(B) Clustered similarity matrix for responses to 73 sounds recorded from a local population. Average population firing rates in response to each sound are shown at the bottom. Examples of the corresponding spectrogram and the spatial response pattern are shown for four sounds.

(C) Same as (B) for a population (same as shown in A) displaying two distinct response modes.

(D) Typical OGB1 calcium signals from 57 simultaneously recorded neurons in an awake passively listening animal.

were displayed in color-coded matrix plots after ordering the sounds with a hierarchical clustering algorithm to reveal potential underlying structures in the space of response patterns (Figures 3B and 3C).

In line with the qualitative observations of high stochasticity (see Figure 2A), we observed that response pattern reliability was generally low. The largest response reliability observed in a population was ranging from 0.62 to 0.02 in unresponsive populations (average: 0.29 ± 0.14 SD, n = 124 local populations, see also Figures 3B and 3C). Despite this high variability, the observed reliability levels were clearly higher than for randomized data sets (Figure S3) demonstrating that specific activity patterns in local populations were indeed present. In a given population, the reliability values formed a continuum between sounds evoking a rather strong response and sounds evoking no response reflecting variations in response probability qualitatively observed in Figure 2A.

When we considered the similarity of responses elicited by different sounds, we observed in the majority of local populations that all reliable responses were highly similar to each other, as indicated by similarity values of the same level as the reliability values (e.g., Figure 3B). In these cases, only a single cluster of sound responses was apparent in the similarity matrix, suggesting that a single type of functional response pattern, or response mode, could be generated in these populations. Interestingly, we also found local populations in which two (Figure 3C) or three clusters of sound responses could be visually identified, indicated by similarity values across the clusters that were much lower than the reliability and intra-cluster similarity values.

We wondered if the presence of only few response patterns may be due the network state induced by the anesthetic. To address this issue we performed a series of experiments in awake, passively listening mice (see Experimental Procedures). We observed that brief sounds evoked population responses in a burst-like manner (Figures 3D and 3E). When constructing clustered similarity matrixes from the response vectors, we observed only a few response modes, similar to the anaesthetized mice (Figures 3F and 3G).

To quantitatively assess the number of sound clusters that could be generated, we developed a statistical test that evaluates the probability that the *N* first major clusters could arise from the randomness of single trial response patterns and the low number of individual sound repetitions rather than reflecting true sound clusters (see Figure S3 and Supplemental Experimental Procedures for details on the implementation). With this test, we could evaluate the maximum number of clusters which gave a statistically significant explanation of the distribution of sound response patterns in a given population. This test was run for 67 populations in which at least two sounds generated response patterns with a reliability level above 0.2. In 74.6% of these populations, the data was best explained by a single response mode, while two or three response modes were detected in 20.9% and 4.5% of the respective populations (Figures 4A and 4B). The same analysis performed on the data from awake, passively listening mice revealed a similar distribution (Figure 4C). This analysis demonstrated that, despite the large number of theoretical response modes that groups of several tens of neurons could generate, local auditory cortex populations generate only a small repertoire of functionally distinct response modes. Interestingly, a similar result was obtained when two second long sounds were presented (Figure S4).

We then sought to determine the spatial organization of the neurons that underlie distinct response modes. We calculated the mean firing rate of neurons in response to the groups of sounds associated to the different modes, which were identified in the above analysis. Interestingly, pairs of response modes observed in a given population corresponded to the firing of partially overlapping subgroups of neurons (Figures 4A and 4D). To assess the similarity of tuning of neurons associated to the same or different subgroups, we computed their signal correlations. We found that members of the same subgroup had significantly higher signal correlations than neuron pairs across groups (same mode: 0.76 ± 0.07, n = 37; different modes: 0.53 ± 0.11, n = 23 modes, Wilcoxon test $p = 2 \times 10^{-4}$). Furthermore, the centroids of the neuronal subgroups corresponding to two distinct response modes were significantly more distant to each other than when the neurons of the local population are spatially randomized (Figure 4E). This indicated an organization of the modes into different spatial domains, which was also visually evident in many examples (Figures 4A and 3C). This observation is consistent with previous estimations of the spatial layout of neurons suggesting a patchy organization of neuronal subgroups in the cortex (Rothschild et al., 2010).

Nonlinear Transitions between Local Response Modes

The low number of observed response modes suggests that local activity patterns form discrete representations of sounds. A prediction from this scenario would be that for a continuous transition between two stimuli exciting two modes an abrupt change in response patterns would be observed because the population could generate no intermediate response pattern. Alternatively, the low number of response modes could merely reflect biases or gaps in the set of tested sounds. To determine if abrupt changes in response patterns could be observed, we first identified local populations in anaesthetized mice showing at least two response modes using a broad set of different sounds (Figure 5A). We selected two basis sounds that were falling in either response mode and constructed linear mixtures from them. Next, we retested the same population with the new set of stimuli to map the transition across modes with higher resolution. When the mixture ratio was varied continuously, we observed abrupt transitions in the population activity patterns

(E) Firing rate estimate for each recorded neuron derived from the traces shown in (D) after temporal deconvolution. Sound stimuli presented at 1 s intervals are indicated by black vertical bars.
(F) Clustered similarity matrix for responses to 34 different short sounds recorded from a local population in an awake passively listening animal. Average population firing rates in response to each sound are shown at the bottom.
(G) Same as (F) for another population displaying two distinct response modes.
See also Figure S3.

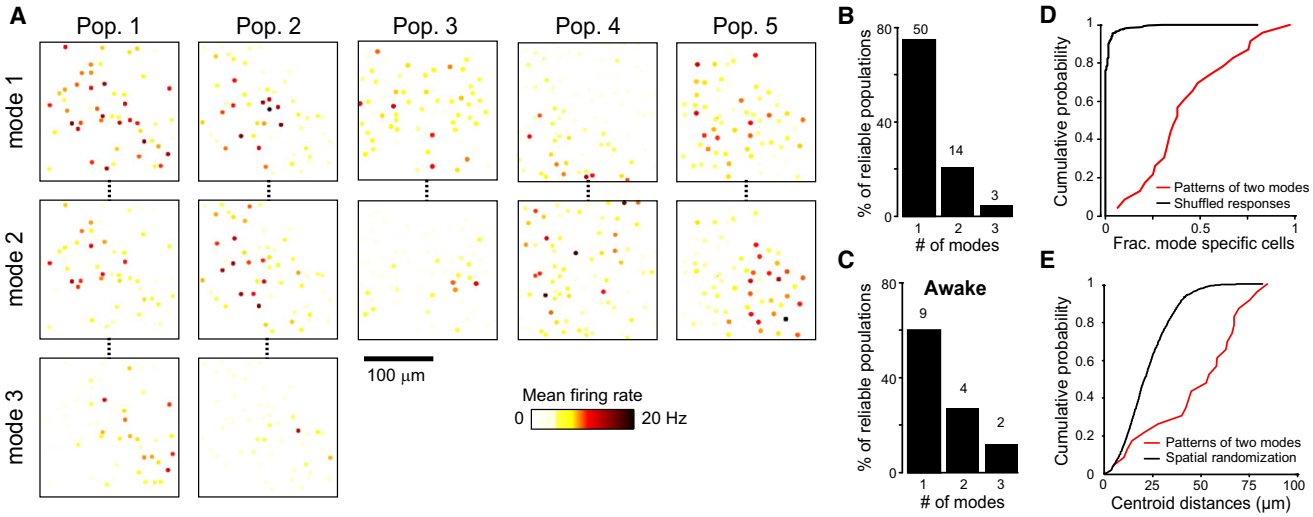

Figure 4. Spatial Distribution of Neurons in Different Response Modes

(A) Examples from five populations (Pop.1, ..., Pop.5) with two to three modes are shown. In these plots, the firing rate is color coded and displayed at the actual 2D location of each neuron to represent the spatial activity pattern. For each neuron, the mean firing rate is computed between 0 and 250 ms after sound onset and averaged across all trials of the sounds falling into a specific mode.

(B) Distribution of the number of detected response modes in all analyzed local populations from anaesthetized mice. Numbers above bars: # of populations.

(C) Same analysis as in (B) performed on data obtained from awake passively listening mice (n = 5).

(D) Cumulative probability plot of the fraction of mode-specific neurons for all pairs of simultaneously recorded response modes (red line). A mode-specific neuron is defined as having significantly different activity levels in each of the modes ($p < 0.01$: Wilcoxon test, comparing the two pooled groups of responses to sounds belonging to each mode). The black line gives the result of the same analysis for populations in which single trials are randomly assigned to a mode (shuffled response).

(E) Cumulative probability distribution of distances between the "centers of mass" of all pairs of response modes observed in the same population (red curve). If the spatial positions of the neurons are shuffled to build more evenly distributed patterns smaller distances are observed (solid black line).

See also Figure S4.

that are visible in both the raw activity plots and the similarity matrices (Figures 5B, 5C, and S5). In addition to the abrupt switch in the structure of the activity patterns, we often observed a decrease in the reliability of one of the response mode close to the actual transition (Figures 5B and 5C). It is noteworthy, that this decrease in reliability can give the impression of a smooth mode transition without a change of the underlying pattern as such. Importantly, in none of the experiments we observed the emergence of a new activity pattern after retesting with a new stimulus set, corroborating the finding of a strong constraint on response modes that are allowed in the network.

To further verify the abrupt change in response pattern we wanted to more carefully assess whether at any point in the linear mixtures both modes might be simultaneously present. Toward this end, we performed the following analysis: we computed the optimal linear decomposition $\vec{R} = \sum_{i=1}^{2} \alpha_i \vec{m}_i + \vec{r}$ over the template patterns $\vec{m}_1$ and $\vec{m}_2$ of the two modes for each single trial response pattern $\vec{R}$ of the sound mixtures. The templates were computed as the average response pattern to each of the two basis sounds, excluding responses to the mixtures. The decomposition was obtained using a standard least square linear fitting algorithm (Moore-Penrose pseudoinverse method) minimizing the norm of the residual pattern $\vec{r}$. In this framework, the coefficients α_1 and α_2 represent the strength of the contribution of each mode in a given single trial response pattern. When we plotted α_1 against α_2 for every single trial response pattern of a given local population (Figure 5D) or for all local populations tested (n = 9; Figure 5E), it became clearly apparent that the two modes did not coexist along the transition. This was indicated by the fact that we did not observe high coefficients for both modes in the very large majority of response patterns. We also observed that the average coefficients α_1 and α_2 were never both much larger than zero for any given sound mixture (Figures 5D–5F). Instead, a clear transition was observed at a certain mixture ratio where the value of at least one of the two coefficients dropped abruptly while the other increased (Figure 5F). To quantify the abruptness of the transition the values of α_1 and α_2 for different mixture ratios where fitted with a sigmoidal function from which we derived the slope at the transition. In all populations tested with linear mixtures of sounds, we observed highly nonlinear transitions, indicated by a maximum slope much larger than 1 for at least one of the modes (n = 9; Figure 5G). When fitting slopes to the average coefficients, it should be kept in mind that a possible modulation of the reliability to elicit a given response patterns with changing mixture ratios can lead to a smoothing of the curve despite the fact that the switch in the structure of the pattern as such is abrupt.

An abrupt switch in response patterns could result from a fast loss of efficiency to evoke the response pattern by the respective component of the mixture. Alternatively, the switch could result from a competition of the two response patterns. In the first

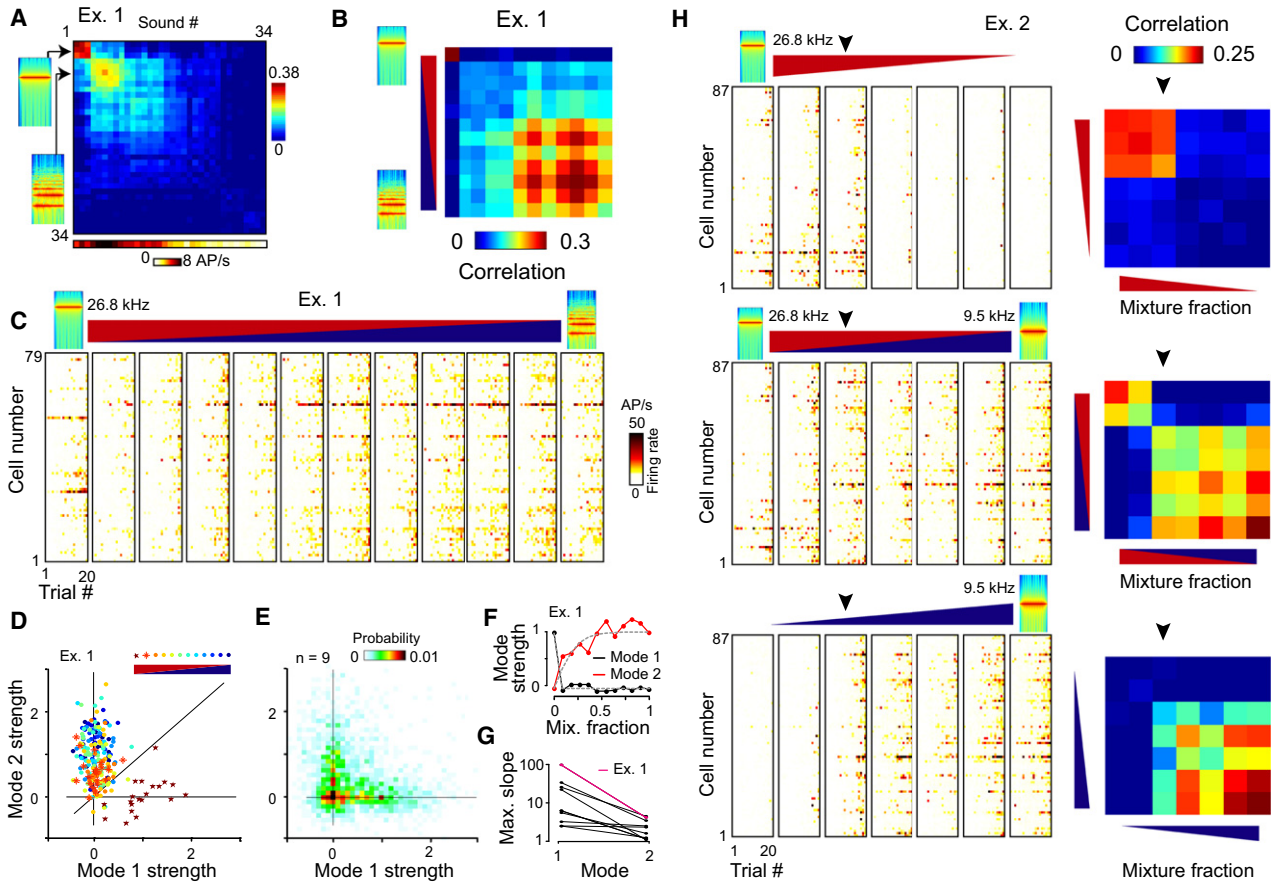

Figure 5. Nonlinear Transitions between Response Modes

(A) Example of a clustered similarity matrix of a population showing two distinct response modes (Ex.1).

(B) Similarity matrix of the same population re-tested with progressive linear mixtures of the two sounds indicated in (A).

(C) Single trial population responses corresponding to the matrices shown in (B) and (C).

(D) Plot of single trial coefficients of the two-mode decomposition for Ex. 1.

(E) Density plot of the single trial coefficients of the two-mode decomposition for the transitions observed in 9 different populations. The color code represents the occurrence probability of a certain combination of the coefficients α_1 and α_2 over all observed transitions. The graph shows that when one mode is strong (high α) the other is suppressed (low α).

(F) Average coefficients of the two-mode decomposition for each mixture along the transition shown in (B). Gray line indicates sigmoidal fit to the data.

(G) Maximal slope of sigmoidal fits (see F) corresponding to the two modes for nine local populations tested. All populations show nonlinear transitions of responses as indicated by slopes much higher than 1.

(H) Single-trial population responses and similarity matrices for different mixtures of two sounds which excited two distinct modes (middle). For this example, the individual components of the mixture were also played independently. The corresponding responses and similarity matrices are shown above and below those of the mixtures. Note that the top-left and middle-left responses as well as the bottom-right and middle-right responses are the same (i.e., sound 1 or 2 played alone).

See also Figure S5.

scenario, the efficacy to elicit a response pattern would be independent of the presence of the second component, whereas in the second scenario the switch would be defined by the specific ratio of the two components making up the mixture. To test this hypothesis, we again identified local populations generating two response modes (n = 5, in 2 mice) and selected two basis sounds exciting each of the modes. We synthesized mixtures of the two basis sounds with seven equally spaced mixture ratios and the individual mixture components in isolation (i.e., one of the basis sounds faded to silence). We compared the patterns of response to the isolated components and to the mixtures and computed the corresponding clustered similarity matrices (Figures 5H and S5). The sound level at which the transition occurred depended on the specific sound and the local population. Importantly, in all cases we found sound levels where both components of the mixture in isolation elicited a reliable response. However, when both components are presented at the same time one of the two response modes appeared to be dominant and only one of the patterns was excited (Figure 5H). Hence, instead of an additive response, the local network falls in a highly nonlinear manner in either one of the two response modes. This indicates that the choice

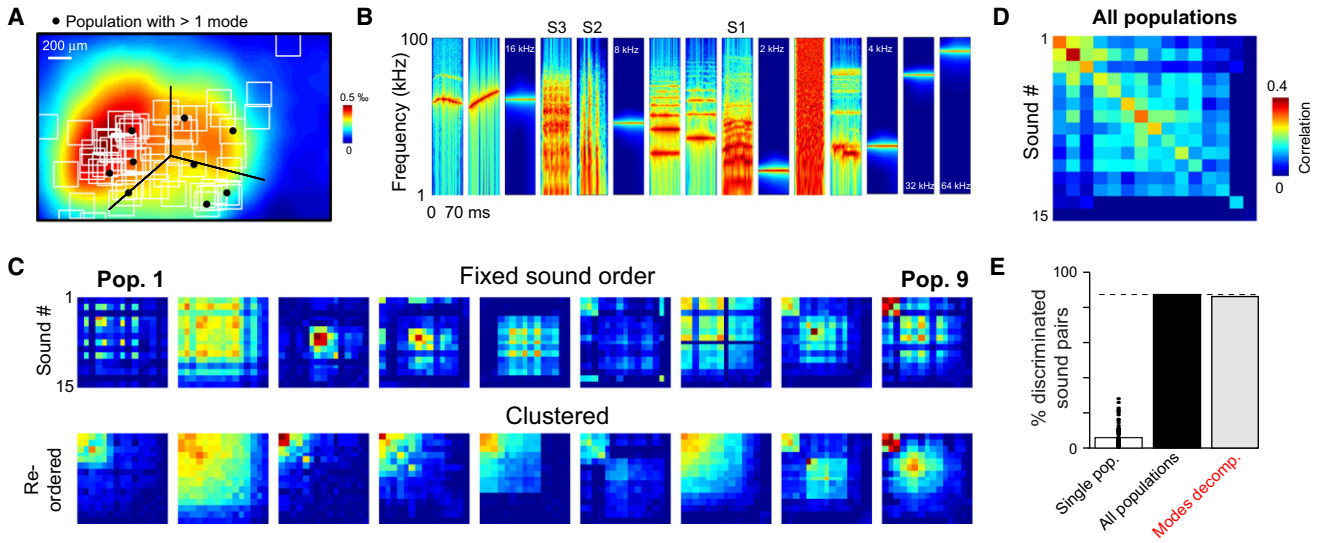

Figure 6. Differential Encoding of Sounds across Local Populations

(A) Average intrinsic signal response to white noise across 12 mice after realignment of the response maps according to the center of A1 tonotopic gradient (see Supplemental Experimental Procedures). The white squares indicate all recording locations in these 12 mice, showing a dense coverage of the primary auditory fields (A1 and AAF). Black dots indicate recording sites from local populations displaying more than one response mode. The black line sketches approximate boundaries between A1 (top left), AAF (top right), and A2 (bottom).

(B) Spectrograms of the sounds ordered according to the similarity of the global population vector (see D). S1, S2, and S3 are the complex sounds used in the discrimination behavior (see Figure 7).

(C) Examples of similarity matrices for the set of sounds shown in (B) from nine different local populations are shown. On the top row, the order of the sounds in the matrix is constant and corresponds to the order of spectrograms shown in (B). On the bottom row, the same matrices are shown, however, the sounds are arranged according to the output of the hierarchical clustering algorithm to show the number of response modes.

(D) Clustered similarity matrix for the same 15 sounds but constructed with the concatenated population vectors of 74 local populations recorded in 14 mice (total of 4,734 neurons). The order of the sounds in this matrix, which results from the clustering of global cortical responses, is as shown in (B). This implies that the sounds which are next to each other in (B) should be similarly perceived according to our global read-out of auditory cortex activity.

(E) Fraction of tested sound pairs (1,770 pairs from 60 sounds) that can be discriminated based on linear classification (SVM) of single-trial response vectors from single local populations with more than 90% success rate (n = 74, white). Same analysis is performed with concatenated full response vectors of the pooled 74 local populations (black), or with the concatenated mode decompositions of the pooled local populations (gray). Mode decomposition gives similar performance despite strongly reduced dimensionality.

See also Figure S6.

of one mode or the other is a winner-take-all decision, which may result from competitive interactions between neuronal populations.

Efficient Sound Encoding with Discrete Representations

Our observations show that local populations of the auditory cortex are constrained to few response modes which encode a small number of sound categories enclosing several sounds. This implies that local populations are highly limited in their capacity to discriminate a large number of sounds. Yet at the level of the organism, sound discrimination does not show such constraints. How this apparent paradox could be resolved became evident when we probed various local populations within and across mice in several primary auditory fields (Figure 6A). In each case, the different local populations categorized different sets of sounds, suggesting that different local populations provide complementary information to unambiguously encode a large number of sounds. To quantitatively assess this observation, we plotted response similarity matrices for a selection of 15 clearly distinct sounds (excluding mixtures and different sound levels; Figure 6B), in which the sound order is fixed (i.e., no clustering was performed; Figure 6C, top). In these plots, the sounds giving rise to a reliable response or being grouped in different modes differ from one population to the next. Thus, different populations are discriminating different sets of sounds. When we built the same similarity matrix for a global AC population using concatenated single-trial population vectors from 74 local populations recorded in 14 anaesthetized mice (total of 4,734 neurons), most of the 15 selected sounds have a higher reliability (diagonal) than their similarity to other sounds (Figure 6D). This indicates that each of the individual sounds can be better discriminated by the global population combining information from multiple local populations.

To precisely quantify general sound discrimination efficiency by local and global populations, we computed the fraction of all possible sound pairs (1,770 pairs from a set of 60 sounds including mixtures and different sound levels) that could be correctly distinguished by a linear support vector machine (SVM) classifier on the basis of single trial population vectors (Shawe-Taylor and Cristianini, 2000). Only 5.7% ± 6.7% (mean ± SD) of sound pairs could be discriminated with greater

than 90% success rate within local populations (Figure 6E). However, up to 87.3% of the sound pairs could be discriminated by the combined activity of our representative sample of AC local populations (74 local populations, total of 4,734 neurons; Figure 6E). Thus, the lack of discrimination at the level of local populations can be overcome by the complementary information provided by the combined activity of the AC. Interestingly, almost identical discrimination efficiency (86.2%; Figure 6E) could be achieved if the dimension of each local population vector $\vec{R}$ was reduced to the number of detected and statistically validated response modes n (1 to 3). This drastic dimensionality reduction was achieved by computing the optimal linear decomposition of $\vec{R}$ over mode templates $\vec{m}_i$ (1 to 3) built as the average of all single-trial population vectors belonging to a given mode as presented above. Note that this method was superior to more standard dimensionality reduction methods such as principal component analysis (Figure S6). This analysis proves that a functional description of the global population activity based solely on the local response modes is sufficient to form a complete representation without loss of information about sounds.

Discrete Representations Predict Behavioral Categorization of Sounds

To what extent do the strongly stochastic and locally discrete representations of sounds in AC actually reflect auditory perception and categorization in the mouse? To answer this question, we tested whether a set of our recordings covering a representative portion of the AC (74 populations, 4,734 neurons in 14 mice, same data set as above) would allows us to quantitatively predict perceptual categorization behavior in mice. We trained mice in a go/no-go task to discriminate a positively reinforced target sound and a negatively reinforced target sound (Figure 7A). We chose three sounds and trained groups of mice to discriminate one of all three pairwise combinations between them. The target sounds included two sounds that our imaging results had shown to elicit more similar patterns and a third sound that elicited a more different pattern. If our imaging results are predictive of behavior, we would expect that the learning rate to discriminate similar patterns is lower that to discriminate more dissimilar patterns. Indeed, in the go/no-go task we observed that the sound pair that elicited the most similar global cortical activity patterns was discriminated with much lower learning rates than the other two pairs. This indicated a correlation between recorded cortical representations and sound discrimination difficulty (Figure 7B), similar to previous reports (Bizley et al., 2010; Engineer et al., 2008).

We observed that mice trained to discriminate a pair of reinforced target sounds would spontaneously react in a consistent manner to other nonreinforced off-target sounds that were presented with a low probability in catch trials. The average response rate to a given off-target sound serves as a report of categorization with respect to the target sounds. This allowed us to obtain a more detailed analysis of the perceived similarity of a broad range of off-target sounds. We observed nonlinear categorization behavior in response to linear mixtures of the two target sounds as indicated by similar response probabilities for a subset of mixtures (Figure 7C). Prediction of spontaneous classification behavior was achieved by a linear support vector machine (SVM) classifier (Shawe-Taylor and Cristianini, 2000) trained to optimally discriminate the single-trial response vectors elicited by the reinforced sounds and tested with vectors elicited by nonreinforced sounds. We observed a good match of the prediction based on global AC activity patterns and behavioral categorization (Figures 7C and 7D; see full results in Figure S7). This match was better than that obtained for alternative descriptions of local population activity using either different time bins for evaluating neuronal firing rates or sequences of time bins to capture some of the information contained in the time course of the response (Figure S7). Interestingly, the best prediction quality was also achieved with the dimensionally reduced description of local activity patterns by mode decomposition (Figure 7E). This demonstrates that the ensemble of local response modes forms a representation that reflects perceived similarity of sounds. In particular, also the nonlinear features of spontaneous categorization behavior were captured.

Discrimination and Categorization of Sounds by Individual Local Populations

We have shown that the nonlinear dynamics of individual local populations spontaneously builds distinct categories of sounds. These sound categories correspond to groups of sounds that excite each of the possible response modes. Also the group of sounds that are unable to elicit a response in a given population can be considered as a category. Could a single local population forming the appropriate categories to distinguish a pair of target sounds be directly used to solve a given discrimination task and would it predict the spontaneous categorization of off-target sounds? To answer this question, we computed for individual local populations the discrimination performance to individual target sound pairs and respective off-target sound categorization. Most populations yielded poor discrimination and poor predictions of sound categorizations (Figures 8A and 8B). However, we observed a few local populations that yielded reliable predictions of categorization behavior for specific target sound pairs comparable to those obtained from the global population vectors (Figure 8A). There were at least 2 to 4 local populations for each target sound pair for which the prediction error was significantly lower than chance levels. The predictive quality of off-target sound categorization by single local populations was correlated with the performance that population in discriminating that target sound pair (Figure 8B). This indicates that neural populations which give the most reliable information to solve the discrimination task readily reflect in their dynamics the behaviorally observed sound categorization (Figure 8C). Therefore, it is conceivable that the sound categories implemented by discrete local response modes form a basis of available perceptual decisions which are selected by learning depending on the behavioral demand.

DISCUSSION

In summary, our findings reveal a coding strategy in the AC in which sound information is distributed globally to counterbalance the limited and stochastic coding observed locally. Our full data set is consistent with classical tonotopic maps;

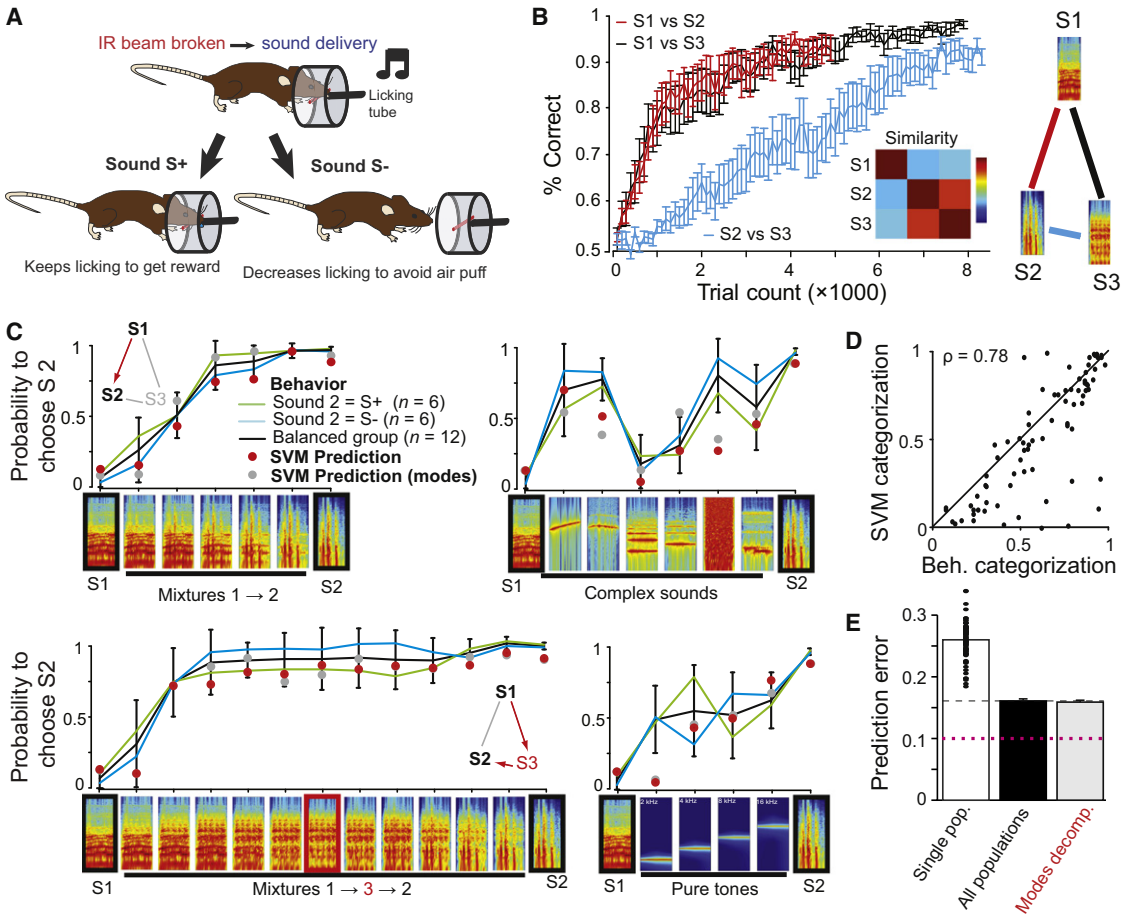

Figure 7. Global Auditory Cortex Representations Predict Categorization Behavior

(A) Schematic of the go/no-go sound discrimination paradigm.

(B) Learning curves (fraction correct trials calculated in blocks of 100 trials) for the 3 pairs of sounds tested for discrimination (n = 12 mice per sound pair, mean ± SEM). The inset shows the similarity matrix (Pearson correlation coefficient) between averaged global population vectors measured for the three target sounds (4,734 neurons, 74 populations, 14 mice). This matrix shows that according to global cortical responses sounds S2 and S3 are expected to be perceived more similar to each other than to sound S1. This observation correlates with the fact that discrimination of S2 and S3 requires a much longer learning than discrimination between S1 and S2 or S1 and S3.

(C) Spontaneous sound categorization for 26 off-target sounds in a group of mice trained to discriminate sound stimuli S1 and S2 (black line, average ± SD, n = 12 mice). The spectrograms of the sounds are shown at the bottom of the graph. (Duration: Pure tones, 50 ms; complex sounds, 70 ms; Frequency range: 1–100 kHz, logarithmic scale.) The colored lines indicate average behavior in the balanced subgroups with switched contingencies (blue: S1 = S+, S2 = S−, n = 6; green: S1 = S−, S2 = S+, n = 6). Red dots correspond to predictions of a linear SVM classifier trained to discriminate concatenated full response vectors of the pooled 74 local populations evoked by S1 and S2 and tested with vectors evoked by off-target sounds (same data set as in Figure 2). Gray dots correspond to analogous predictions of a linear SVM classifier based on concatenated response vectors whose dimensionality had been reduced by mode decomposition (see text).

(D) Behavioral sound categorization versus SVM classifier prediction for the three target sound pairs and all off-target sounds tested. The black line is the identity line. ρ: Pearson correlation coefficient.

(E) Average absolute prediction error of the linear SVM classifier for single-trial population vectors from single local populations (white), or concatenated full response vectors of the pooled 74 local populations (black), or with the concatenated mode decompositions of the pooled local populations (gray). Again, mode decomposition gives similar performance despite strongly reduced dimensionality. The behavioral replicate (red line) is measured as the mean absolute discrepancy between measurements made on two balanced groups of 6 mice. Error bars: SD.

See also Figure S7.

however, the discreteness of local network response patterns was unexpected, since it was widely assumed that AC neurons build a continuum of receptive fields even at local scales. Our observations provide direct evidence that the auditory cortex network is constituted of partially overlapping subnetworks in which individual neurons play redundant roles as recently proposed in an earlier study to explain the spatial distribution of pairwise correlations (Rothschild et al., 2010). This has the important implication that the smooth shape of trial-averaged single cell tuning curves largely reflects variations in the probability to elicit the same, stereotyped stochastic network pattern. Our recordings were performed in a 200 × 200 μm field of view.

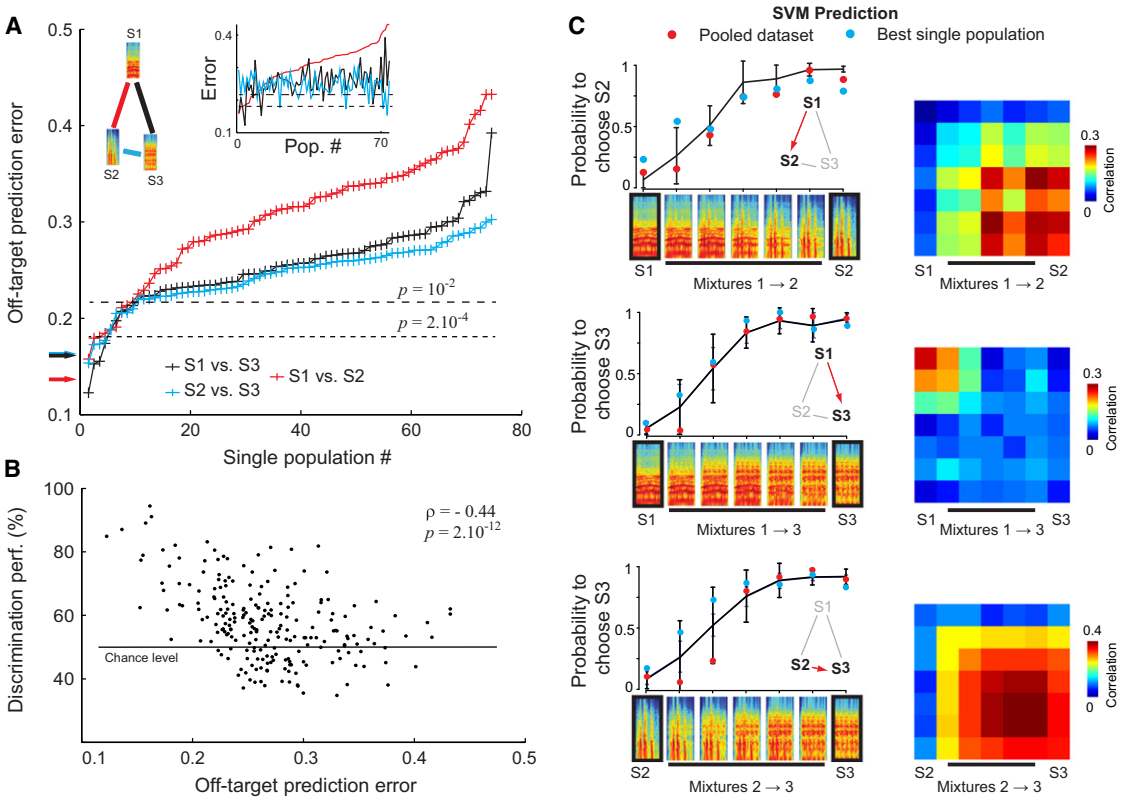

Figure 8. Discrimination and Categorization of Sounds by Individual Local Populations

(A) Mean error in the prediction of off-target sound categorization behavior for each of the 74 individual local populations (crosses) and for the global population (arrows). Each of the three target sound pairs trained for discrimination are tested independently (red, black, blue). Populations are ordered according to the mean prediction error for each target sound pair. The inset shows the same plots but with the same population order for the three target sound pairs. Significance values were derived from the distribution of prediction errors obtained after 10,000 randomizations of the off-target sound labels in the behavioral data. 1% level indicated by the dashed line; 5% significance level divided by the Bonferroni correction for multiple testing indicated by the dotted line.

(B) Mean error in the prediction of off-target sound categorization behavior plotted against target discrimination performance for individual local AC populations and sound pair discriminations. Low prediction errors correlate with high discrimination performance (correlation coefficient: $\rho = -0.44$, p-value, $p = 2.10^{-12}$, Student's t distribution transformation method).

(C) (Left) Prediction of the categorization of linear mixtures of the target sounds by the local population discriminating best the targets (different local population for each target pair). Error bars: SD. (Right) Clustered similarity matrix for the population responses to the sound mixtures in the local population shown on the left.

The fact that almost 80% of them showed a single response mode could indicate that the typical spatial extent of the sub-networks corresponding to a response mode is significantly larger. While our observations are consistent with a columnar organization of the mouse auditory cortex (Mountcastle, 1997), it should be noted that the dynamics of the infragranular layers is to some extent dissociated from the dynamics of layers II and III and thus the organization of sound evoked patterns in discrete response modes could be a specificity of the supragranular layers (Sakata and Harris, 2009).

One important result of our study is that the network activity carries little information about sounds at the local scale because of the high constraint on local activity patterns. However, since different local populations are responsive to different groups of sounds, the combination of response patterns from various local ensembles leads to a significant increase in encoding capabilities. Therefore, our results suggest that a complete representation of sounds emerges only at a global scale, potentially encompassing whole auditory fields. This is in line with the observation that arrays of local field potential recordings in the human brain are sufficient to retrieve much of the information about sounds despite their lack of spatial precision and their inability to account for single-neuron activity (Pasley et al., 2012). Thus, our results corroborate the idea that the function of the auditory cortex is dominated by broad scale dynamics in which groups of hundreds of neurons rather than single neurons form the functional units. Our results also demonstrate that these functional units are capable of producing discrete network states.

The discrete nature of local response modes is highlighted by the fact that when two modes are observed at the same location they cannot simultaneously coexist, and transitions between these between the two response modes are highly nonlinear (Figure 5). This observation is further corroborated by the comparison of response patterns elicited by mixed sound stimuli and by their individual components which reveals that

subnetworks corresponding to two modes interact in a competitive fashion (see Figures 5H and S5; Kurt et al., 2008). This dynamics could result of large scale excitatory and in particular inhibitory interactions generating collective, attractor-like dynamics (Hopfield, 1982) and may be an optimal strategy to encode information under noisy conditions (Tkacik et al., 2010).

Based on theoretical considerations, discrete network states have been proposed to underlie the formation of categories and objects in brain circuits (Hopfield, 1982). However, it is only recently that experimental evidence for such dynamics has been obtained (Niessing and Friedrich, 2010; Wills et al., 2005). Our observation that, in naive mice, various novel sounds trigger the same local response pattern indicates that local AC ensembles spontaneously form categories of stimuli independent of prior training to specific sounds. We show that discrete representations measured by calcium imaging quantitatively reflect spontaneous categorization of sounds measured in a behavioral task. Therefore, the discrete network dynamics in AC are compatible with behaviorally measured perceptual categorization. Predictions of the categorization behavior are based on a linear classifier (SVM). This classifier is mathematically equivalent to a binary neuron that would be linearly summing up inputs from the recorded auditory cortex neurons (Shawe-Taylor and Cristianini, 2000), similar to the perceptron model (Rosenblatt, 1958). Interestingly, such a simple architecture allowed us to predict the behavioral response even when mice alternated between two choices with close to 50% probability (Figure 7D). This implies that AC activity reflected not only the direction of the categorization but also large parts of the observed variability in the categorization behavior, suggesting that behavioral variability could be explained in principle to a large extent from the variability of the sensory representation itself.

While we think that much of the sound information is distributed globally, it is remarkable that a good prediction of the categorization behavior and of its variability could even be obtained with well-chosen single local populations (Figure 8). This indicates that the perceptual decisions made by the animal can be parsimoniously explained by the selection of a relatively small group of neurons that spontaneously provides a suitable categorization of sound stimuli. Such a model would be distinct from scenarios in which learning leads to an optimal adaptation of a plastic decision boundary to a continuous sensory representation.

What may be the functional role of discrete dynamics in circuits of the auditory cortex? In the primate visual system, complex category signals (objects or groups of objects) are classically reported in higher areas such as inferior temporal, parietal, and prefrontal areas (DiCarlo et al., 2012; Swaminathan and Freedman, 2012). The observation of category-forming dynamics already in a primary sensory area suggests that this may be a general property of neocortical circuits. It is conceivable that higher-order categories are built on a hierarchy of lower-order categories which arise in primary sensory areas. Such a hierarchical structure of discrete representations might be essential for elaborate cognitive functions such as language processing. The fact that, e.g., phonemes are perceived and thereby stably recognized as discrete sound categories (Liberman et al., 1967) might rely on similar dynamics of the human auditory cortex.

EXPERIMENTAL PROCEDURES

Animals

Experimental subjects were male CB57BL/6J mice (Charles River, age: 6 to 16 weeks) and were performed in accordance with the Austrian laboratory animal law guidelines and approved by the Viennese Magistratsabteilung MA58 (Approval #: M58/02182/2007/11; M58/02063/2008/8).

Sounds

All sounds were delivered free field at 192 kHz sampling rate in a sound proof booth by a custom-made system consisting of a linear amplifier and a ribbon loudspeaker placed 25 cm from the mouse head (Audiocomm, Vienna, Austria). The transfer function between the loudspeaker and the location of the mouse hear was measured using a probe microphone (4939-L-002, Brüel&Kjær, Bremen, Germany) and compensated numerically by filtering the sound files with the inverse transfer function to obtain a flat frequency response at the mouse ear (between 0.5 kHz and 64 kHz ± 4 dB). Sound control and equalization was performed by a custom Matlab program running on a standard personal computer equipped with a Lynx 22 sound card (Lynx Studio Technology, Inc, Costa Mesa, CA).

Sounds were 50 ms pure tones of various frequencies or 70 ms complex sounds characterized by broad frequency content and temporal modulations, generated from arbitrary samples of music pieces or animal calls played with fourfold speed. All stimulus onsets and offsets were smoothed with a 10 ms long half-period cosine function. Series of K linear sound mixtures were generated as $Mixture(k) = (1\text{-}k/K)\ Sound_1 + k/K\ Sound_2$ with $0 \leq k \leq K$. The first set of stimuli tested in imaging experiments contained 60 sounds: 2, 4, 8, 16, 32, 64 kHz pure tones at 4 intensities (50 to 80 dB SPL at 10 dB interval), 3 broadband complex sounds at 5 intensities (53 to 81 dB SPL at 7 dB interval), 6 broadband complex sounds at 74 dB SPL and 15 mixtures of the first 3 complex sounds (74 dB SPL) and was used in 14 mice for prediction of behavioral sound categorization (Figures 6, 7, and 8). In two experiments, 73 sounds were played including additionally 46 dB SPL complex sounds, 40 dB SPL pure tones and one more mixture series (examples in Figure 3). A third set of stimuli contained 34 sounds covering a broader range of spectrotemporal parameters: 19 pure tones (2 to 45 kHz) and 15 broadband complex sounds (74 dB) and was used in 10 mice for studying transitions between modes (Figure 5) and for testing the linear prediction of complex sound responses (Figure S2). This set of stimuli was also used in 5 mice for awake experiments (Figures 3D–3G and 4C). The statistical determination of the number of modes in local populations (Figure 4) was run on experiments in which the sets of either 60, 73, or 34 sounds were used.

Intrinsic Imaging

To determine the location of calcium imaging recordings with respect to the functional organization of auditory fields, we routinely performed intrinsic imaging experiments under isoflurane anesthesia (1%), a day after calcium imaging. The brain was incidentally illuminated through the cranial window by a red (intrinsic signal: wavelength = 780 nm) or a green (blood vessel pattern: wavelength = 525 nm) LED. Reflected light was collected at 25 Hz by a CCD camera (CCD1200QD, Vosskuehler GmbH, Germany) attached to a macroscope consisting of two objectives placed face-to-face (Nikon 135 mm and 50 mm; Soucy et al., 2009). The focal plane was placed 400 μm below superficial blood vessels. A custom-made Matlab program controlled image acquisition and sound delivery. We acquired a baseline and a response image (170 × 213 pixels, ~3.1 × 2.4 mm) as the average illumination image 2 s before and 2 s after sound onset, respectively. For each trial, the change in light reflectance (ΔR/R) was computed as (baseline − response)/baseline (note that with this convention increase in brain activity translates into positive ΔR/R values). For each sound, 30 trials were acquired, averaged and low-pass filtered (Gaussian kernel, σ = 5 pixels) to build the response map. Sounds were trains of 20 white noise bursts or pure tone pips (80 ms—2, 4, 8, 16, 32 kHz) separated by 20 ms smooth gaps.

In Vivo Calcium Imaging

A craniotomy (~1 × 2 mm) was performed above the right auditory cortex under isoflurane anesthesia (1.5% to 2%). A standard solution (Garaschuk

et al., 2006) containing 1 mM of the calcium-sensitive dye Oregon Green BAPTA 1 (OGB1) was pressure ejected (0.7 bar: 10 pulses of 10 s) at ~10 loci of the auditory cortex region through a thin glass pipette (~5 MΩ tip resistance). The craniotomy was closed with a thin cover glass, sealed with dental cement (Ortho-Jet, Lang Dental, Wheeling, IL). After a 30 min recovery period, the animal was head-fixed under the imaging apparatus and kept under isoflurane anesthesia (1%). Fields of neurons in cortical layers 2/3 (~150–300 μm below the dura) were imaged using a two-photon microscope (Ultima IV, Prairie Technologies, Middleton, WI) equipped with a 20x objective (XLUMPlan FI, n.a. = 0.95, Olympus, Tokyo, Japan). OGB1 was excited at 950 nm using a pulsed laser (Chameleon Ultra, Coherent). Line scans (33 to 25 lines/s) over visually selected neurons were used to record OGB1 fluorescence changes (see also Supplemental Experimental Procedures for details on the line scan design). For a given recording site, imaging was performed in less than 30 min. We did not observe a significant change in sound-evoked firing rates during this period (2.9 ± 0.1 AP/s, SD over the 15 trials, ANOVA, p = 0.39, n = 74 populations).

Calcium Imaging in Awake Mice

Mice habituated to head-fixation underwent OGB-1 injection and window implantation following procedures used in a previous report and described in detail in the Supplemental Experimental Procedures (Komiyama et al., 2010). To allow off-line compensation of movement artifacts images were acquired in full-frame mode (128 × 128 pixels, 162.2 ms sampling interval). Deconvolution of calcium traces and construction of clustered similarity matrices was performed as for data from anaesthetized mice.

Cell-Attached Recordings

To establish a relationship between the observed changes in fluorescence and the actual firing rate of a neuron, we performed in a number of experiments simultaneous calcium imaging and cell-attached recordings. Cell-attached recordings were obtained with pulled, thin wall glass pipettes (5 to 8 MΩ tip resistance) filled with intracellular solution (in mM: 130 K-gluconate, 5 KCl, 2.5 $MgCl_2$, 10 HEPES, 0.6 EGTA, 4 Na_2ATP, 0.4 Na_3GTP, 10 Na_2-phosphocreatine, and 0.03 sulforhodamine for visualization). Extracellular voltage was amplified by an ELC-03XS amplifier (NPI, Tamm, Germany) and digitized through a Digidata1440A (Molecular Devices). We recorded action potentials elicited by sounds or by ejection of currents (up to 100 nA) through the recording pipette.

Deconvolution of Calcium Signals

All recordings consisted of blocks of 10–15 s separated by more than 2 s. To evaluate the baseline fluorescence F_0, the onsets t_i of calcium transients were detected as peaks of the first derivative of the raw signal that were two standard deviations above the mean. F_0 was obtained by fitting the linear model $F_0 + \sum_i a_i \theta(t - t_i) \exp(-(t - t_i)/\tau)$ to the raw fluorescence signal $F(t)$ (θ is a step function and τ = 1.3 s) using the Moore-Penrose pseudoinverse. After having obtained an estimate of F_0 the normalized change in calcium fluorescence was computed as $\Delta F/F = (F(t)-F_0)/F_0$. The neuronal firing rate $r(t)$ was evaluated by deconvolution of the normalized fluorescence change $\Delta F/F = (F(t)-F_0)/F_0$:

$$r(t) = \alpha\left(\frac{d\Delta F/F}{dt} - \frac{1}{\tau}\Delta F/F\right).$$

τ = 1.3 s and α = 0.018 are the typical decay time and $\Delta F/F$ amplitude of the calcium transient triggered by a single action potential. They were determined to minimize the error between estimated and actual firing rate observed in simultaneous in vivo cell attached recordings and imaging.

Response Pattern Analysis

For a local population of N neurons recorded simultaneously the response pattern to presentation i of sound p was represented by the population vector $\vec{R}_{p,i} = (r_{k,p,i})_{k\in[1:N]}$ of dimension N where each entry contains the firing rate of one of the N neurons averaged between 0 and 250 ms following sound onset (note that other time bins were also analyzed; see Figure S7). We defined the response similarity between sounds p and q as $S_{p,q} = 1/M^2 \sum_{i=1}^{M}\sum_{j=1}^{M} \rho(\vec{R}_{p,i}, \vec{R}_{q,j})$ with $\rho(\vec{A}, \vec{B})$ being the Pearson correlation coefficient between $\vec{A}$ and $\vec{B}$, and M being the number of presentations of a sound. This corresponds to the average correlation of all possible pairwise combinations of single trial response vectors of two sounds. Similarly, we defined the reliability of the response to sound p as $S_{p,p} = 2/(M^2 - M)\sum_{i=1}^{M}\sum_{j=i+1}^{M} \rho(\vec{R}_{p,i}, \vec{R}_{p,j})$. In all displayed matrices, sounds were sorted using the standard single link agglomerative hierarchical clustering algorithm implemented in Matlab to group sounds that elicit similar response patterns. The statistical method to determine the number of significant clusters is described in Supplemental Experimental Procedures.

The distance between the "centers of mass" of the mean response patterns corresponding to two modes was computed as

$$d = \left\| \sum_{k\in[1:N]} \left(\frac{r_{k,\text{mode 1}}}{\sum_{k\in[1:N]} r_{k,\text{mode 1}}} - \frac{r_{k,\text{mode 2}}}{\sum_{k\in[1:N]} r_{k,\text{mode 2}}} \right) \begin{pmatrix} x_k \\ y_k \end{pmatrix} \right\|$$

where x_k and y_k are the two-dimensional spatial coordinates of neuron k in the field of view $r_{k,\text{ mode 1}}$ and $r_{k,\text{ mode 2}}$ are the mean firing rates of this neuron in each response mode. The "center of mass" of a response pattern can be viewed as the average position of most active neurons in the pattern. The signal correlation between a pair of neuron was computed as the Pearson correlation coefficient between the two vectors containing the average firing rate responses (250 ms time bin starting at sound onset) of each of the neurons for all sounds tested in the particular experiment. Signal correlations were computed for mode-specific neurons associated to the same mode or to different modes. A mode-specific neuron is defined as having significantly higher activity levels in one of the modes of the local population ($p < 0.01$: Wilcoxon test, comparing the pooled groups of responses to sounds belonging to each mode).

Sound Discrimination and Categorization by Cortical Representations

The efficiency of single trial sound pair discrimination by cortical representations was evaluated using linear Support Vector Machine classifiers (Shawe-Taylor and Cristianini, 2000) implemented in Matlab using the OSU SVM toolbox (http://sourceforge.net/projects/svm/). Half of the single trial population vectors were used as training set to determine the maximum margin classifier between vectors representing each sound. This classifier was then tested with the remaining trials to compute the fraction of correctly classified trials. To predict behavioral categorization, the linear classifier optimized to distinguish the cortical responses to the two target sounds of the behavioral discrimination task (1 and 2) was tested with single trial response patterns evoked by off-target sounds. The fraction of trials classified as sound 1 (or 2) gave our estimate of the probability of choosing the response appropriate for sound 1 (or 2). For both sets of analysis, we used alternatively local population vectors containing the responses of a set of neurons recorded simultaneously or global population vectors consisting of the concatenated populations vectors (in full or reduced by mode decomposition) from several local populations and mice.

Sound Discrimination and Categorization Behavior

Water deprived mice were trained daily in a 30 min session of ~200 trials to obtain water reward (~5 μl) by licking on a spout over a threshold after a positive target sound S+ and to avoid a 10 s air puff by decreasing licking below this threshold after a nonrewarded, negative target sound S−. Both sounds consisted of two 4 kHz pips (50 ms) followed after a 375 ms interval by a specific 70 ms complex sound taken from the set of sounds used for imaging. Licking was assessed 0.58 s after the specific sound cue in a 1 s long window by an infrared beam system which detected the presence of the mouse's snout immediately next to the licking spout (Coulbourn instruments, PA). The licking threshold was set to be 75% beam-break duration in the assessment window. Sound delivery and valve control for water reward and air puff was performed by a custom Matlab program. Positive and negative sounds were played in a pseudorandom order with the constraint that exactly 4 positive and 4 negative sounds must be played every 8 trials.

Performance was measured as the fraction of correct positive and correct negative trials over all trials. Once a mouse had reached at least 80% correct performance, 1 of 27 off-target sounds (26 sounds + 1 blank off-target) randomly replaced a target sound in one over 10 trials followed by no reinforcement. In a given session only 9 out of 27 off-target sounds were presented. Given two target sounds, 1 and 2, spontaneous categorization of off-target sounds was measured as the probability that the mouse makes the correct response for sound 2 after hearing a specific off-target sound. We observed that categorization measurements beyond the 8 first trials started to display a small systematic drift. This drift could result from learning that off-target sounds which are categorized as the positively reinforced sound in fact do not yield a reward. Therefore, we included only the 8 first trials of any given off-target sound for our estimation of categorization behavior.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2012.07.008.

ACKNOWLEDGMENTS

We thank M. Palfreyman, A. Zador, and Y. Loewenstein for comments on the manuscript, A. Helm, A. Bichl, M. Ziegler, and M. Colombini for technical assistance, and T. Wernle and G. Loevinsohn for pilot experiments. This work was supported by Boehringer Ingelheim GmbH and a postdoctoral fellowship to B.B. from the Human Frontier Science Program. B.B. performed imaging experiments, behavioral experiments, and analysis of the data. L.U. performed behavioral experiments. B.B. and S.R. designed research and wrote the manuscript.

Accepted: July 3, 2012
Published: October 17, 2012

REFERENCES

Amit, D.J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cereb. Cortex *7*, 237–252.

Atiani, S., Elhilali, M., David, S.V., Fritz, J.B., and Shamma, S.A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron *61*, 467–480.

Bandyopadhyay, S., Shamma, S.A., and Kanold, P.O. (2010). Dichotomy of functional organization in the mouse auditory cortex. Nat. Neurosci. *13*, 361–368.

Bartho, P., Curto, C., Luczak, A., Marguet, S.L., and Harris, K.D. (2009). Population coding of tone stimuli in auditory cortex: dynamic rate vector analysis. Eur. J. Neurosci. *30*, 1767–1778.

Bathellier, B., Buhl, D.L., Accolla, R., and Carleton, A. (2008). Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. Neuron *57*, 586–598.

Bizley, J.K., Walker, K.M., King, A.J., and Schnupp, J.W. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. J. Neurosci. *30*, 5078–5091.

Christianson, G.B., Sahani, M., and Linden, J.F. (2008). The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. J. Neurosci. *28*, 446–455.

Christianson, G.B., Sahani, M., and Linden, J.F. (2011). Depth-dependent temporal response properties in core auditory cortex. J. Neurosci. *31*, 12837–12848.

David, S.V., Mesgarani, N., Fritz, J.B., and Shamma, S.A. (2009). Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. J. Neurosci. *29*, 3374–3386.

DeWeese, M.R., and Zador, A.M. (2006). Non-Gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. J. Neurosci. *26*, 12206–12218.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434.

Eggermont, J.J. (2011). Context dependence of spectro-temporal receptive fields with implications for neural coding. Hear. Res. *271*, 123–132.

Engineer, C.T., Perez, C.A., Chen, Y.H., Carraway, R.S., Reed, A.C., Shetake, J.A., Jakkamsetti, V., Chang, K.Q., and Kilgard, M.P. (2008). Cortical activity patterns predict speech discrimination ability. Nat. Neurosci. *11*, 603–608.

Garaschuk, O., Milos, R.I., and Konnerth, A. (2006). Targeted bulk-loading of fluorescent indicators for two-photon brain imaging in vivo. Nat. Protoc. *1*, 380–386.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA *79*, 2554–2558.

Hromádka, T., Deweese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol. *6*, e16.

Kalatsky, V.A., Polley, D.B., Merzenich, M.M., Schreiner, C.E., and Stryker, M.P. (2005). Fine functional organization of auditory cortex revealed by Fourier optical imaging. Proc. Natl. Acad. Sci. USA *102*, 13325–13330.

Komiyama, T., Sato, T.R., O'Connor, D.H., Zhang, Y.X., Huber, D., Hooks, B.M., Gabitto, M., and Svoboda, K. (2010). Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature *464*, 1182–1186.

Kurt, S., Deutscher, A., Crook, J.M., Ohl, F.W., Budinger, E., Moeller, C.K., Scheich, H., and Schulze, H. (2008). Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. PLoS One *3*, e1735.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). Perception of the speech code. Psychol. Rev. *74*, 431–461.

Loebel, A., Nelken, I., and Tsodyks, M. (2007). Processing of sounds by population spikes in a model of primary auditory cortex. Front. Neurosci. *1*, 197–209.

Luczak, A., Barthó, P., and Harris, K.D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron *62*, 413–425.

Maass, W., Joshi, P., and Sontag, E.D. (2007). Computational aspects of feedback in neural circuits. PLoS Comput. Biol. *3*, e165.

Machens, C.K., Wehr, M.S., and Zador, A.M. (2004). Linearity of cortical receptive fields measured with natural sounds. J. Neurosci. *24*, 1089–1100.

Miller, E.K., Nieder, A., Freedman, D.J., and Wallis, J.D. (2003). Neural correlates of categories and concepts. Curr. Opin. Neurobiol. *13*, 198–203.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science *319*, 1543–1546.

Mountcastle, V.B. (1997). The columnar organization of the neocortex. Brain *120*, 701–722.

Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. Nature *397*, 154–157.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? Biol. Cybern. *89*, 397–406.

Niessing, J., and Friedrich, R.W. (2010). Olfactory pattern classification by discrete neuronal network states. Nature *465*, 47–52.

Ohl, F.W., Scheich, H., and Freeman, W.J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. Nature *412*, 733–736.

Pasley, B.N., David, S.V., Mesgarani, N., Flinker, A., Shamma, S.A., Crone, N.E., Knight, R.T., and Chang, E.F. (2012). Reconstructing speech from human auditory cortex. PLoS Biol. *10*, e1001251.

Rabinowitz, N.C., Willmore, B.D., Schnupp, J.W., and King, A.J. (2011). Contrast gain control in auditory cortex. Neuron *70*, 1178–1191.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. *65*, 386–408.

Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. Nat. Neurosci. *13*, 353–360.

Russ, B.E., Lee, Y.S., and Cohen, Y.E. (2007). Neural and behavioral correlates of auditory categorization. Hear. Res. *229*, 204–212.

Sakata, S., and Harris, K.D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron *64*, 404–418.

Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. Annu. Rev. Neurosci. *33*, 203–219.

Shawe-Taylor, J., and Cristianini, N. (2000). Support Vector Machines (Cambridge: Cambridge University Press).

Soucy, E.R., Albeanu, D.F., Fantana, A.L., Murthy, V.N., and Meister, M. (2009). Precision and diversity in an odor map on the olfactory bulb. Nat. Neurosci. *12*, 210–220.

Swaminathan, S.K., and Freedman, D.J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat. Neurosci. *15*, 315–320.

Tkacik, G., Prentice, J.S., Balasubramanian, V., and Schneidman, E. (2010). Optimal population coding by noisy spiking neurons. Proc. Natl. Acad. Sci. USA *107*, 14419–14424.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. J. Neurosci. *24*, 10440–10453.

Wang, X.J. (2008). Decision making in recurrent neuronal circuits. Neuron *60*, 215–234.

Wills, T.J., Lever, C., Cacucci, F., Burgess, N., and O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. Science *308*, 873–876.

Yaksi, E., and Friedrich, R.W. (2006). Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. Nat. Methods *3*, 377–383.