

# Subthreshold membrane responses underlying sparse spiking to natural vocal signals in auditory cortex

Krista E. Perks<sup>1</sup> and Timothy Q. Gentner<sup>1,2,3,4</sup>

<sup>1</sup>Neurosciences Graduate Program, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, USA

<sup>2</sup>Department of Psychology, University of California San Diego, La Jolla, CA, USA

<sup>3</sup>Neurobiology Section, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

<sup>4</sup>Kavli Institute for Brain and Mind, La Jolla, CA, USA

**Keywords:** auditory cortex, hierarchical models, natural stimuli, object selectivity, synaptic integration

## Abstract

Natural acoustic communication signals, such as speech, are typically high-dimensional with a wide range of co-varying spectro-temporal features at multiple timescales. The synaptic and network mechanisms for encoding these complex signals are largely unknown. We are investigating these mechanisms in high-level sensory regions of the songbird auditory forebrain, where single neurons show sparse, object-selective spiking responses to conspecific songs. Using whole-cell *in vivo* patch clamp techniques in the caudal mesopallium and the caudal nidopallium of starlings, we examine song-driven subthreshold and spiking activity. We find that both the subthreshold and the spiking activity are reliable (i.e. the same song drives a similar response each time it is presented) and specific (i.e. responses to different songs are distinct). Surprisingly, however, the reliability and specificity of the subthreshold response was uniformly high regardless of when the cell spiked, even for song stimuli that drove no spikes. We conclude that despite a selective and sparse spiking response, high-level auditory cortical neurons are under continuous, non-selective, stimulus-specific synaptic control. To investigate the role of local network inhibition in this synaptic control, we then recorded extracellularly while pharmacologically blocking local GABAergic transmission. This manipulation modulated the strength and the reliability of stimulus-driven spiking, consistent with a role for local inhibition in regulating the reliability of network activity and the stimulus specificity of the subthreshold response in single cells. We discuss these results in the context of underlying computations that could generate sparse, stimulus-selective spiking responses, and models for hierarchical pooling.

## Introduction

Stimulus encoding – the relationship between an external event and the accompanying neural response – is the cornerstone of sensory neurophysiology (Adrian, 1926). Yet, for the complex sensory signals that are essential to many natural behaviors, our understanding of stimulus encoding is poor. In particular, we know very little about the synaptic inputs evoked by natural signals, and the operations governing their integration and transformation into spiking responses in single neurons.

Here we test specific predictions about the selectivity of stimulus-specific synaptic drive underlying sparse, selective spiking in the auditory cortex of European starlings, a species of songbird. Songbirds, in particular starlings, are well suited for these studies. Starling songs are acoustically complex and composed of very diverse, brief segments (motifs) that are perceived as distinct auditory objects (Gentner, 2008). Stimulus-driven spiking activity in the higher-order cortical regions caudal mesopallium (CM) and caudo-medial nidopallium (NCM) is sparse: only a small portion of all possible motifs evoke robust spiking in single neurons (Gentner &

Margoliash, 2003; Meliza *et al.*, 2010; Thompson *et al.*, 2010; Jeanne *et al.*, 2011; Meliza & Margoliash, 2012) and each motif evokes spiking from only a small number of neurons (Gentner & Margoliash, 2003). Responses to song elements are dependent on acoustic context (Jeanne *et al.*, 2011; Kozlov & Gentner, 2014), and (in zebra finches) combining song elements into longer bouts increases the sparseness of spiking responses (Schneider & Woolley, 2013).

Sparse spiking responses to natural signals appear to be a general property of sensory cortex (Vinje & Gallant, 2000, 2002; Olshausen & Field, 2004; Graham & Field, 2007b; Hromadka *et al.*, 2008; Sakata & Harris, 2009; Tolhurst *et al.*, 2009). That is, only a small proportion of all stimuli evoke spikes from any given neuron (lifetime sparseness, which we refer to as selectivity), and only a small proportion of neurons spike at any point in time (population sparseness). Sparse representations convey a range of computational benefits to downstream neurons for the classification and recognition of complex signals (Ganguli & Sompolinsky, 2012; Babadi & Sompolinsky, 2014). Models for object recognition and classification rely on feed-forward hierarchical pooling of the outputs from simpler feature detectors to build sparse, selective spiking responses to increasingly complex signals that carry behaviorally relevant information (Riesenhuber

Correspondence: Dr T. Q. Gentner, <sup>1</sup>Neurosciences Graduate Program, as above.  
E-mail: tgentner@ucsd.edu

Received 12 September 2014, revised 7 December 2014, accepted 11 December 2014

& Poggio, 1999). The synaptic and network mechanisms underlying this pooling remain unclear.

We distinguish two potential synaptic pooling regimes: 'sparse' and 'distributed'. Each regime makes different predictions about the selectivity of the subthreshold response underlying selective spiking in single neurons. In a sparse pooling regime, neurons with selective spiking responses pool inputs that are necessarily less selective but still biased towards the features in complex natural signals that ultimately drive changes in spike rates. Sparse synaptic pooling predicts that stimulus-specific subthreshold activity will be selective (i.e. driven only by a subset of the potential stimuli), and that the degree of selectivity will vary only quantitatively from that of the spiking response. Moreover, the timing of stimulus-specific subthreshold responses should, on average, align with the stimulus-evoked changes in spike rate. We contrast sparse pooling with distributed pooling, in which neurons with selective spiking responses pool synaptic inputs that are not biased towards the features in complex natural signals that drive changes in spike rates. Distributed pooling predicts that the net stimulus-evoked synaptic input will be stimulus-specific (i.e. unique for different stimuli), but non-selective (i.e. driven by all stimuli). Because it is non-selective, the stimulus-specific subthreshold activity under distributed pooling will not necessarily be tied to stimulus-evoked spike rate changes. So long as the inputs are stimulus-specific, however, the cell could be made to spike selectively (through a variety of mechanisms) to a potentially wide range of relevant features.

The foregoing hypothetical pooling scenarios both predict that the subthreshold membrane response of neurons whose stimulus-driven spiking responses are selective will contain stimulus-specific activity. Only the sparse pooling hypothesis requires that the stimulus-specific subthreshold activity is selective for subsets of stimuli. To distinguish these two hypothetical pooling scenarios, we compare properties of the subthreshold response to spiking. Direct measurement of the subthreshold activity along with spiking was accomplished using whole-cell, *in vivo* recording techniques in NCM and CM of starlings during presentation of natural conspecific vocalizations, i.e. songs. We characterized variability across time and across trials for both subthreshold and spiking activity. In concert with the observation of very selective (sparse) stimulus-driven spiking responses, we find strong, remarkably persistent, stimulus specificity in the subthreshold response to all songs regardless of spiking. These results provide strong evidence against sparse synaptic pooling and support instead a distributed synaptic pooling regime in which, in sharp contrast to spiking output, net subthreshold responses are stimulus-specific and non-selective.

In a distributed pooling regime, stimulus-specificity throughout the subthreshold response depends on the trial-to-trial reliability of the inputs. Inhibition is a widespread feature of cortical networks (Wilent & Contreras, 2005; Ayaz & Chance, 2009; Isaacson & Scanziani, 2011) and in Starling NCM, local inhibition modulates the selectivity of spiking responses (Thompson *et al.*, 2013). We test whether inhibition is necessary for maintaining stimulus specificity in the subthreshold response throughout stimulation. If inhibition modulates the reliability (and by extension the specificity) of the synaptic input to individual neurons then it would necessarily modulate the reliability of the network spiking response, and the subthreshold and spiking responses of individual neurons. We find that transient local blockade of gamma-aminobutyric acid (GABA) receptors decreases the reliability of stimulus-driven spiking responses in NCM, consistent with a role of inhibition for supporting a non-selective, stimulus-specific distributed synaptic pooling regime underlying selective spiking in NCM.

## Methods

### Animal preparation

Experiments used adult European starlings (*Sturnus vulgaris*), wild-caught in southern California. We prepared the starlings for the recording session by attaching a small pin stereotaxically to the surface of the skull with dental cement (under isoflurane anesthesia). For electrophysiological recordings, we anesthetized the starlings with 20% urethane (7–8 mL/kg, in three to four intramuscular injections over ~2 h) prior to being placed in the recording chamber or with a continuous infusion of 1.3% ketamine in 5% glucose saline at 2 mL/kg/h throughout the recording. Starlings were placed in a cloth jacket and secured via the attached pin to a stereotaxic apparatus inside a sound attenuation chamber. A small craniotomy was made dorsal to the recording site (NCM: 0–300 rostral of Y sinus and 500 lateral of midline; CM: 2500 caudal of Y sinus and 500–1500 lateral of midline), the dura removed and electrodes advanced into the brain.

### Ethical standards

All procedures were conducted in accordance with approved IACUC protocols and in accordance with the guidelines laid down by the NIH in the US regarding the care and use of animals for experimental procedures.

### Electrophysiology

*Whole cell patch* current clamp recordings (MultiClamp 700B amplifier; Axon Instruments, Union City, CA, USA) of 5 to 60-min duration were made using the blind patch technique (Margrie *et al.*, 2002). Whole-cell patch pipettes (3 to 6-M $\Omega$  tip resistance) were fabricated from filament (0.25 mm) borosilicate glass (OD 2.0 mm, ID 1.5 mm; Hilgenberg, Malsfeld, Germany). The standard K<sup>+</sup>-based internal solution was: (in mM) potassium gluconate 135, NaCl 8, HEPES 10, Mg-ATP 4, Na-GTP 0.3, EGTA 0.3 (pH 7.4, 298 osm). Recordings were obtained by slowly advancing the electrode through the region of interest (about 1500–2000  $\mu$ m below the surface) while monitoring its resistance with voltage steps. During the initial descent through the hyperpallium, a large amount of positive pressure (~300 mbar) was applied to keep the electrode tip free from debris. After arriving at the depth of interest, the positive pressure was reduced to 25–35 mbar, capacitance compensation was adjusted, and the pipette was advanced in 2- to 3- $\mu$ m steps until direct contact with the cell membrane was detected as an increase in resistance. Immediately upon contact, pressure was released and formation of the giga-seal between the electrode tip and the cell membrane occurred either spontaneously or after slight suction applied by mouth. After the giga-seal stabilized (typically within a few minutes) suction was used to obtain whole cell access (access resistances ranging from 5 to 90 M $\Omega$ ; see 'Intrinsic physiology').

### Extracellular recordings

To examine the role of local inhibition in controlling spiking reliability in the network, we used data from extracellular recordings in NCM collected as a part of previous experiments (Thompson *et al.*, 2013). Briefly, commercial multibarreled glass pipettes containing a carbon fiber electrode (5  $\mu$ m diameter; 400–1200 k $\Omega$  impedance) and six attached barrels (~3  $\mu$ m diameter) were used for drug microiontophoresis (Kation Scientific, Minneapolis, MN, USA). Gabazine (SR95531, 3 mM, pH 3.2; Sigma Aldrich, St Louis, MO,

USA), or a gabazine/saclofen combination, was used to inhibit GABA-mediated inhibition locally in NCM around the recording site. The combined application of gabazine and saclofen ( $n = 12$  sites) did not elicit responses different from those during application of gabazine alone (Thompson *et al.*, 2013). Following Thompson *et al.* (2013) we considered drug delivery at a particular site to be successful if the average song-driven firing rate during iontophoresis was significantly different (either higher or lower) from the firing rate prior to iontophoresis.

### Auditory stimulation

All stimuli were extracted from previously recorded song repertoires of adult European starlings. Single motifs (stereotyped multi-note elements of natural Starling song) played two or three times in succession, or longer segments (5–10 s) of continuous song were played to the anesthetized animal in an anechoic recording chamber. Auditory stimuli were presented free field from a full-range speaker mounted 30 cm from the center point of the subject's head, where the mean sound pressure level ranged from 40 to 80 dB SPL.

### Data analysis

Electrical activity recorded in whole-cell configuration was low-pass filtered (10 kHz), digitally sampled (44.1 kHz), and saved for offline analysis (IGOR PRO2; WaveMetrics Software, Lake Oswego, OR, USA). For further analysis, data were down-sampled to 10 kHz and exported to a format used by custom-written MATLAB (Mathworks Software) routines. Only stimulus–response pairs for which there were at least five repeats were included in the analyses of the whole-cell data.

### Intrinsic physiology

We present data recorded in whole-cell configuration from 20 single neurons in NCM and CM. As this is the first report using whole-cell *in vivo* recordings in CM and NCM of the starling, we include here a basic characterization of intrinsic properties for reference/comparison with other brain regions and species. To estimate the passive input resistance ( $R_{in}$ ) and time constant ( $T$ ) of the membrane, we applied a negative current pulse (–75 pA) through the recording pipette and fit the voltage response with a double exponential function to isolate the electrode artifact from the membrane response. To avoid contamination by slow, voltage-activated conductance in our estimate of the passive membrane properties ( $R_{in}$  and  $T$ ), we fit only the first 100 ms of the voltage response. The median series resistance ( $R_s$ ) across all 20 neurons used in this study was 32 M $\Omega$  (CI: 5–64 M $\Omega$ ), which allowed sufficient isolation of the membrane response. The median  $R_{in}$  was 157 M $\Omega$  (CI: 56–309 M $\Omega$ ) and the median  $T$  was 14 ms (CI: 7 and 21 ms). Independence of estimates for  $R_{in}$  and  $T$  from  $R_s$  was confirmed with a simple linear regression model.

The negative current step also regularly induced slowly depolarizing voltage sags. We estimated the time course and magnitude of this effect by removing the passive components of the voltage response (by subtracting the exponentials used to estimate  $R_{in}$  and  $R_s$ ), and then re-fitting the remaining voltage response with a single exponential (Zhu *et al.*, 1999). In all cases there was a significant exponential fit with a median time constant of 184 ms (CI: 110–364 ms) that was depolarizing in 19/20 cases (across which the steady-state membrane voltages ranged from –96 to –64 mV). In a subset of neurons recorded for another experiment, we measured the

dependence of the input resistance on the membrane potential and found a positive correlation. This has not been well characterized, but is consistent with hyperpolarization-induced activation of a conductance.

In 14 of 20 neurons, the series resistance was sufficiently low (< 50 M $\Omega$ ) to analyse the temporal dynamics of spike shape. For these neurons, the median width at half-height was 1.0 ms (CI: 0.5–1.7 ms) and the median spike threshold was 25 mV above the resting membrane potential (CI: 19–30 mV).  $R_s$  did not correlate with spike height.

We measured a limited number of properties but the data did not suggest any clear distinction into cell types. The reported analyses therefore pool across all recorded cells in two brain regions. Sampling biases inherent in whole-cell patch techniques (relating to factors such as soma size, morphology and myelination) are probably present in our dataset; other cell types not recorded could, of course, show very different responses. Notwithstanding such issues, the consistency of our main results of the study across a potentially diverse set of cell types highlights the robustness of the effects.

### Signal filtering

To detect spike times from intracellular records, the signal was high-pass filtered at 200 Hz and thresholded. For extracellularly recorded data using carbon-fiber electrodes, spike times were detected as reported by Thompson *et al.* (2013). Measures of sub-threshold activity were made after the voltage record was smoothed using a one-dimensional, 18-ms median filter to suppress noise and clip spikes near threshold.

### Response epoch detection

One important component of our experiment is the use of natural stimuli. As such, we did not want to arbitrarily tailor the duration of our stimulus to a duration that was experimentally convenient to analyse a 'response', as is common when using artificial stimuli. Instead, under the assumption that the activity across the entire duration of a complex stimulus does not necessarily constitute a meaningful response, we presented long bouts of song within which we defined multiple shorter responses. To do this, we separately parsed the ongoing spiking and subthreshold activity using a simple algorithm that identified localized responses (relative to the pre-stimulus trial-averaged activity), which we call 'response epochs'. To detect significant subthreshold response epochs we took the median-filtered voltage response to a song stimulus and overlaid a sliding 200-sample (20 ms)-wide analysis window broken into 20 bins. At each time-step the activity within that window was considered a response if 85% of the bins had values that exceeded the confidence bounds set by the trial-averaged pre-stimulus activity (Fig. 1A). We detected significant spiking response epochs in the same way, except that we first created a smoothed spiking probability function by converting the vector of raw spike times to a binary string where '1' indicated a spike and '0' no spike, convolving with a narrow Gaussian filter, averaging across trials, and then normalizing to max = 1. To parse spiking activity we used a 500-sample (50 ms)-wide analysis window broken into 50 bins and 95% confidence intervals on the pre-stimulus activity. To maximize the number of independent responses and minimize the duration of responses, it was necessary to use slightly different values for the window size and confidence bounds in the analysis of spiking and subthreshold activity. Although the statistics of the subthreshold activity vector necessitated a smaller analysis window than for spiking, the minimum

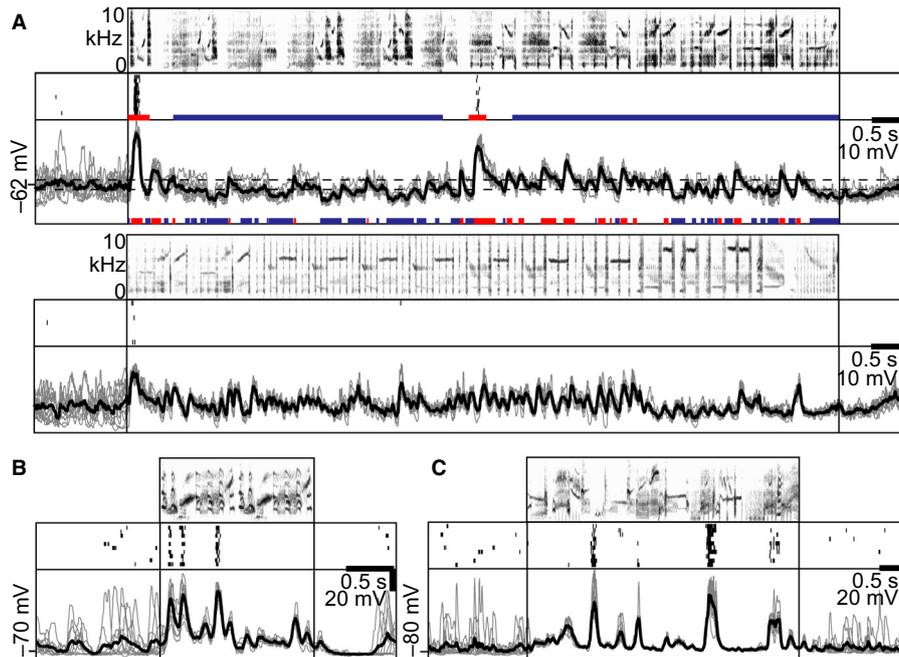


FIG. 1. Examples of spiking and subthreshold activity recorded in whole-cell configuration. (A) Neural recording from an example neuron before, during and after the presentation of two different 10-s starling song segments sampled from a longer bout. Top, spectrogram of the song. Middle, raster of spike times for each of 10 trials in which the song sample above was presented. Bottom, isolated subthreshold membrane potential recordings (gray) for the same neuron on the same 10 trials as above, overlaid with the trial-averaged voltage in black. Colored bars below spiking and subthreshold activity denote facilitatory (red) and suppressive (blue) response epochs (see Methods); mean trial-averaged pre-stimulus voltage =  $-62$  mV. Dotted lines mark the 98% confidence bounds on the pre-stimulus range of trial-averaged subthreshold activity. (B,C) Example spiking and subthreshold activity recorded from two different neurons across trials during which song segments sampled from longer bouts were played. Neural activity and song are displayed as described in (A).

response duration was smaller than the analysis window for both spiking and subthreshold responses (62 and 22 ms, respectively) indicating that the width of the analysis window did not set the lower bound on distributions of response epoch durations in either case and that comparisons can be made between the results obtained in both activity regimes.

#### Stimulus specificity of subthreshold potential

We recorded the membrane potential throughout the presentation of each song stimulus across multiple trials (5–10 trials) and used a median filter to clip spikes. To assess the stimulus specificity of subthreshold activity (whether the variance in the membrane potential within a given portion of song across trials is smaller than the variance between different portions) we used k-means clustering. Offline, we segmented the full song and the corresponding time-locked subthreshold activity into evenly spaced bins of a given duration (durations used for the analysis ranged from 0.09 ms to 0.9 s). We refer to each binned song segment as a ‘stimulus’ for our analysis. We applied k-means clustering to a random subset of five bins and sorted all subthreshold activity (which we refer to as ‘response’ for our analysis) within that set of bins based on similarity in (i) the time-averaged membrane potential and (ii) the temporal pattern of membrane voltages. To measure clustering accuracy we calculated the proportion of correct response–stimulus assignments contained in the k-means result for every possible bin–cluster permutation and took the maximum proportion correct as the clustering accuracy. We iterated this process 100 times and calculated the average clustering accuracy for each cell–stimulus pair and each segment duration. The noise floor for the clustering accuracy was calculated using the subthreshold activity recorded during silence before song onset.

#### Fano factor analysis

We adapted routines for analysing mean spike rate and Fano factor on the extracellular data set from the ‘variance toolbox’ available from the Churchland lab and used in a recent report (Churchland *et al.*, 2010). The mean-matched Fano factor was computed for all stimulus conditions across the set of extracellular sites using a 50-ms sliding analysis window (in 25-ms steps) to provide a time-varying estimate of the reliability of the spike rate throughout stimulation. Mean matching equalizes the firing rate distributions across time to control for the dependence of Fano factor on firing rate.

#### Statistical analysis

All data were tested for normality using the Kolmogorov–Smirnov test evaluated at  $P < 0.05$ . When appropriate, central tendencies are reported as median  $\pm$  the 95% confidence interval calculated from the cumulative distribution function unless otherwise stated. Non-parametric tests were used when data were not normal.

#### Results

In this study we distinguish two hypothetical synaptic pooling strategies that could both support hierarchical object selectivity, but that make distinct predictions about the selectivity of stimulus-specific subthreshold activity underlying selective spiking in single neurons in high-level auditory cortex. To examine spiking and subthreshold responses, we recorded neural activity using whole-cell patch clamp techniques in 20 single neurons in regions NCM ( $n = 12$ ) and CM ( $n = 8$ ) of anesthetized starlings presented with a range of conspecific songs. These songs have a wide range of natural variation in

the distribution of spectro-temporal features across time. This time-varying acoustic structure is reflected in the high temporal variability of both the spike rate and the subthreshold activity (Fig. 1). Although both the spiking and the subthreshold response vary considerably across the duration of the stimulus, as shown in the example neurons (Fig. 1), they are nonetheless quite reliable for each repetition of the same stimulus. In characterizing both the within-trial variability and the between-trial reliability over multiple time-scales, we use the term ‘time-averaged’ to refer to the mean activity averaged over the duration of a single stimulus presentation, and the term ‘trial-averaged’ to refer to the time-varying activity averaged across multiple presentations of the same stimulus.

#### Variability in spiking and subthreshold activity across time

We first quantified the within-trial variability of both the spiking and the subthreshold activity in individual neurons. For spiking activity, we isolated epochs during stimulation for which the trial-averaged spike probability density function exceeded the trial-averaged pre-stimulus range (see Methods). We sorted these epochs into facilitating or suppressing responses depending on whether the time-averaged spike rate increased or decreased relative to the pre-stimulus period (Figs 1A and 2A; see Methods). The duration of these spiking responses varied (median 217 ms, CI = 62–1122,  $n = 174$  facilitatory responses; median 386 ms, CI = 75–2854,  $n = 116$  suppressive responses) as did their time-averaged spike rate (median 11.1 spikes/s, CI = 0.4–33.4,  $n = 174$  facilitatory responses; median 0 spikes/s, CI = 0–0,  $n = 116$  suppressive responses). On average, the facilitatory responses constituted 8% of the total stimulus duration (CI = 0–55%;  $n = 116$  stimuli), and suppressive responses constituted 9% of the stimulus duration (CI = 0–82%;  $n = 116$  stimuli). We note that the low spontaneous spike rates observed (median 1.0 spike/s, CI = 0.3–3.3,  $n = 20$  neurons) are common in these regions (Gentner & Margoliash, 2003; Keller &

Hahnloser, 2009; Schneider & Woolley, 2013) and can make suppression of spiking difficult to measure. Thus, the actual number of suppressive responses may be greater.

For subthreshold activity, we isolated response epochs during song presentation in which the trial-averaged voltage exceeded the trial-averaged pre-stimulus range (see Methods). We sorted these epochs into facilitating or suppressive responses based on whether the time-averaged voltage within each epoch was depolarized or hyperpolarized relative to the pre-stimulus mean (Fig. 1A; see Methods). These subthreshold responses varied in duration (median 83 ms, CI = 26–449,  $n = 638$  facilitatory responses; median 61 ms, CI = 22–302,  $n = 332$  suppressive responses) and the time-averaged polarization (relative to the minimum potential recorded during silence; median +11 mV, CI = +7 to +19,  $n = 638$  facilitatory responses; median +2.8 mV, CI = +0.9 to +4.1,  $n = 332$  suppressive responses). In a few cases where we calculated the resting membrane potential in voltage clamp (data not shown), we found that the minimum membrane potential recorded during silence approximated it. Across all neurons in our sample, the facilitatory subthreshold responses constituted on average 30% of the total stimulus duration (CI = 2–80%;  $n = 116$  stimuli), which was a significantly larger proportion than the 8% of the stimulus that contained facilitatory spiking responses (see above;  $P < 0.0001$  Kolmogorov–Smirnov test,  $n = 116$  stimuli).

Spiking always co-occurs with subthreshold depolarization (facilitatory response), and facilitatory spiking responses are by definition a subset of facilitatory subthreshold responses. Therefore, to estimate the amount of facilitatory subthreshold responses that did not co-occur with spiking responses we take the difference between the total proportion of the stimulus in which there is facilitatory subthreshold response (30%) and subtract the total proportion in which there is a facilitatory spiking response (8%). Thus, approximately 22% of facilitatory subthreshold responses occur in the absence of spiking. Based on membrane voltage alone, only a small fraction

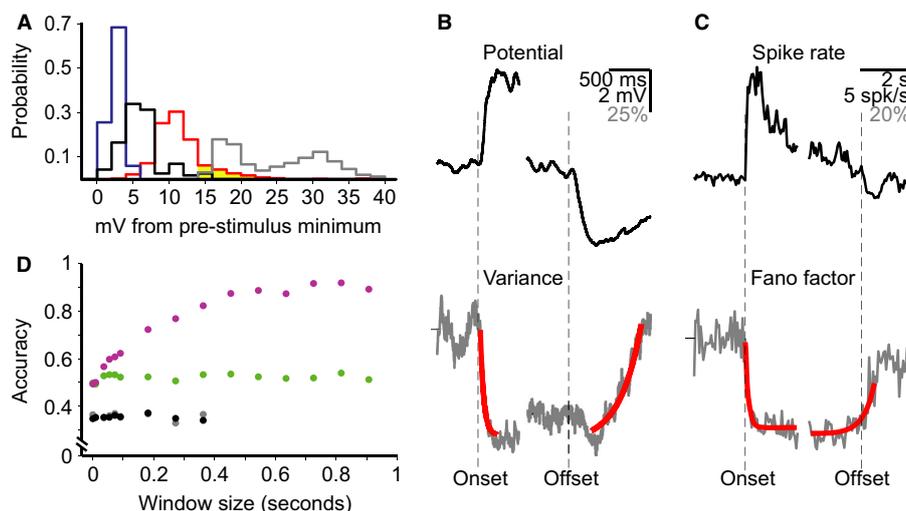


FIG. 2. Stimulus-driven response reliability and specificity. (A) Histograms of mean subthreshold activity level preceding each stimulus (black), the threshold potential for all spikes recorded from all neurons (gray), and membrane potential values contained within all facilitatory (red) and suppressive (blue) response epochs. (B) The mean trial-averaged, filtered membrane potential (top) and the median trial-to-trial variance relative to pre-stimulus silence (bottom; hash = median variance before stimulus onset) across all neurons and stimulus blocks. (C) The mean spike rate (top) and the mean-matched Fano factor relative to pre-stimulus silence (bottom; hash = median Fano factor before stimulus onset) computed for all stimulus conditions across the set of extracellular sites using a 50-ms sliding window. Data in (B) and (C) are aligned to stimulus onset and offset (stippled lines). (D) Mean accuracy (from k-means clustering) with which any given segment of the membrane response can be correctly distinguished from any other segment of similar duration, for segments ranging from 0.09 ms to 0.9 s. Different colored points show the classification accuracy for the membrane voltage time series (magenta), the time-averaged membrane potential (green) during stimulus presentations, along with the same measures made during the silent interval preceding stimulus onset (gray, mean; black, time-series) for each duration of analysis time-window.

(~17%) of the membrane potential values contained within facilitatory subthreshold responses exceeded the minimum spike threshold (16 mV above the pre-stimulus minimum voltage; Fig. 2A), and thus had the potential to elicit a spike. The actual distribution of spikes within this subset of voltages will be lower, as it depends on the refractory period and the local voltage derivative. Regardless of the exact spiking distribution, robust spiking responses are much more selective than the facilitatory subthreshold activity.

Suppressive subthreshold responses were more rare than facilitatory subthreshold responses (2% stimulus duration, CI = 0–42%;  $n = 116$  stimuli), and more rare than suppressive spiking (9% stimulus duration, CI = 0–82%;  $n = 116$  stimuli). Of course, spiking activity can be suppressed by subthreshold responses with a range of mean voltages anywhere below spike threshold.

### Reproducibility of spiking and subthreshold activity

The two synaptic pooling strategies we address in this study are distinguished in part by the extent of stimulus control over the synaptic activity. Across-trial variability in the subthreshold and spiking activity is a reliable metric of stimulus control that we can easily measure throughout stimulation. The foregoing analyses (which showed that both spiking and subthreshold responses span a range of magnitudes and durations; Figs 1 and 2A) imply that the activity within response epochs was reproducible across trials. Reliability is not necessarily constrained to those epochs and may be measurable on a smaller timescale. We therefore measured reliability of subthreshold and spiking activity throughout stimulation by calculating the trial-to-trial variance for the subthreshold activity (single sample resolution) and the Fano factor for spiking (25-ms resolution) and characterized their temporal dynamics at stimulus onset and offset.

We observed a significant decrease in the trial-to-trial variance of the membrane potential during stimulation compared with the pre-stimulus period (Fig. 2B, bottom;  $P < 0.0001$  Kolmogorov–Smirnov test; median decrease relative to baseline = 68%, CI = 84–25,  $n = 115$  cell–stimulus pairs). The time course for the onset of this drop in the membrane potential variance was very rapid ( $\tau = 41$  ms), but the relaxation back to the pre-stimulus levels at the offset of the stimulus was much slower ( $\tau = 262$  ms) (Fig. 2B, bottom). The suppressed variance relative to baseline persisted throughout song stimulation, but for individual response epochs the variance depended on the mean as follows. For facilitating responses (depolarized potentials relative to pre-stimulus mean) the variance of the membrane potential decreased by 85% relative to baseline ( $n = 20$  cells). For suppressing responses (hyperpolarized potentials relative to the pre-stimulus mean) the variance of the membrane potential decreased by 99.8% relative to baseline ( $n = 20$  cells). We note that this difference is likely to be influenced by several factors, including voltage-dependent non-linearities that increase the spiking probability for depolarized responses, and differences in the number of synaptic events contributing to each response.

We quantified trial-to-trial variability in the spiking activity before, during, and after stimulus presentation using the Fano factor with mean-matching techniques to control for firing rate dependencies (Fig. 2C; Churchland *et al.*, 2010). This analysis is data-intensive and requires more trials than were recorded under whole-cell configuration and more neurons in order to correct for firing rate biases. To measure the between-trial reproducibility of spiking activity, we used extracellular multi-unit activity recorded in NCM to song stimuli (five adult starlings, 42 multi-unit sites; 336 site–stimulus pairs; Thompson *et al.*, 2013). At stimulus onset there was a

rapid, significant decrease in the Fano factor ( $\tau = \sim 119$  ms; single exponential fit), to 70% of baseline (CI = 68–81,  $P < 0.0001$  Kolmogorov–Smirnov test,  $n = 336$  stimuli). At stimulus offset, the Fano factor estimate relaxed back toward the pre-stimulus level, but with a much slower time course than the onset ( $\tau = \sim 412$  ms; Fig. 2C bottom).

### Stimulus specificity of subthreshold fluctuations across multiple timescales

The reproducibility of subthreshold responses suggests that neurons are under continuous synaptic control throughout the stimulus, rather than intermittently receiving input corresponding only to features most relevant to changes in output spike rates. By itself, however, trial-to-trial reliability does not confer stimulus specificity. Spiking responses in CM and NCM are both selective (show lifetime sparseness) and stimulus-specific (respond distinctly to different stimuli) (Jeanne *et al.*, 2011). Having demonstrated that the subthreshold activity is reliable, we next determined whether it is also stimulus specific throughout song presentation. Specifically, does the same portion of a song produce a unique membrane voltage response (temporal pattern or mean value) that is similar across stimulus repetitions?

To answer this question in our whole-cell data set, we applied k-means analysis to calculate an average clustering accuracy across response–stimulus pairs (see Methods: ‘Stimulus specificity of subthreshold potential’). We then took this accuracy as a measure of subthreshold stimulus specificity for each cell–stimulus pair (Fig. 2D). We measured stimulus specificity using a range of stimulus durations, but always included data recorded during the entire song bout in the analysis. The noise floor for the accuracy estimate was calculated using the activity recorded during the silence before the stimulus came on. Both the mean and temporal pattern of membrane potential activity contained enough stimulus specificity to allow for accurate clustering of response–stimulus pairs (compared with the ‘noise’ floor) even at single-sample resolution. Beginning at relatively short timescales (~40 ms), the temporal pattern of subthreshold activity could be clustered more accurately than the mean (Fig. 2D). Thus, even very short, randomly-chosen portions of the membrane activity carry stimulus-specific information. Increasing the size of the analysis window substantially improved clustering accuracy for temporal patterns, but not for the mean, demonstrating that the time-varying membrane potential carries additional stimulus-specific information.

### Inhibition modulates across-trial reproducibility of spiking in NCM

We reasoned that local inhibition might also mediate the robustness of stimulus specificity in the subthreshold activity, which depends on the reliability of spiking activity in the network from which its inputs are pooled. To test this idea, we blocked GABA receptors transiently around extracellular recording sites using iontophoretic application of gabazine (see Methods).

Local inhibition modulates the magnitude of neuronal spiking in NCM across song stimuli (Thompson *et al.*, 2013), but its effect on individuated response epochs was not previously tested. We parsed spiking response epochs (as in Fig. 1A) using the activity recorded in the gabazine condition and compared those response epochs with the corresponding stimulus–response epochs in the intact condition. Blocking local inhibition induced an increase in spiking activity during silence (median = 1.5 spikes/s, CI = 0–15.0, intact condition;

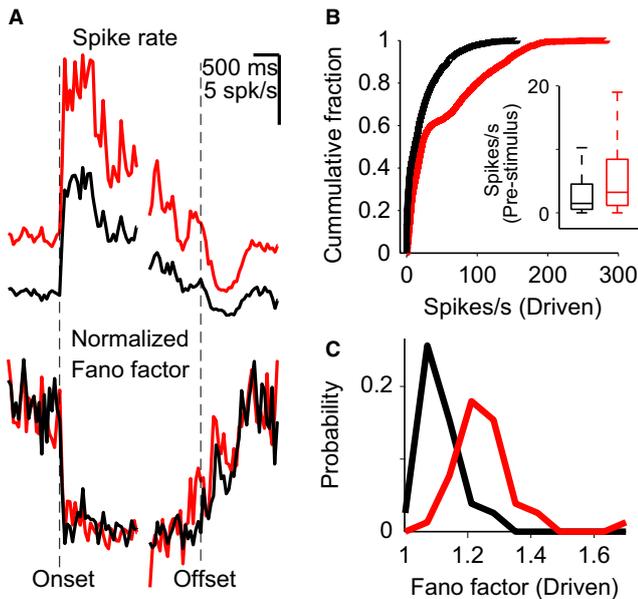


FIG. 3. Effects of gabazine. (A) The spike rate (top) and the mean-matched Fano factor (bottom) during the intact condition (black) and during gabazine iontophoresis (red), computed for all stimulus conditions across the set of extracellular sites using a 50-ms sliding analysis window (in 25-ms steps). For visualization of the relative onset and offset time course, the Fano factor trace in each condition is aligned to the median stimulus-evoked Fano factor and normalized to the maximum Fano factor before stimulus onset. (B) Cumulative distribution functions of the spike rate per response epoch (colors as in A). Inset: boxplot of spike rates during the pre-stimulus period in each condition (horizontal line shows the median; the box shows the 25th and 75th percentiles; the whiskers encompass all non-outlier data: ~99.3%). (C) Histogram of Fano factor estimates for stimulus-driven activity (colors as in A).

median = 3.2 spikes/s, CI = 0.2–24.8, gabazine condition;  $n = 296$  stimuli) and a shift toward response epochs with higher spike rates during auditory stimulation (median = 11.6 spikes/s, CI = 0–71.6, intact condition; median = 21.5 spikes/s, CI = 1.8–162.2, gabazine condition;  $n = 1815$  responses) consistent with previously reported effects (Thompson *et al.*, 2013; Fig. 3A and B).

The extracellularly-recorded spiking activity (from single and multi-unit sites) can serve as a reasonable proxy for estimating the reliability of the local synaptic network driving a randomly chosen cell within that network. To test whether local inhibition might contribute to the reliability of the local network activity, we compared the Fano factor computed across stimulus trials (see Methods) when local inhibition was intact and when it was blocked by gabazine. Indeed, the distribution of stimulus-driven Fano factor estimates is shifted significantly to larger values when local inhibition is blocked by gabazine than when inhibition is intact (median = 1.13 intact and 1.27 gabazine;  $P < 0.0001$ , Kolmogorov–Smirnov test; Fig. 3C). Notably, the Fano factor distribution across sites during the pre-stimulus period was not significantly altered by gabazine (median = 1.52 intact and 1.53 gabazine;  $P = 0.7$ , Kolmogorov–Smirnov test). Likewise, the time course of the change in variance at stimulus onset and offset appeared to be unaffected by blocking local inhibition with the application of gabazine ( $\text{Tau}_{\text{onset}} = 119$  ms and  $\text{Tau}_{\text{offset}} = 412$  ms with inhibition intact;  $\text{Tau}_{\text{onset}} = 115$  ms and  $\text{Tau}_{\text{offset}} = 358$  ms with inhibition blocked; Fig. 3A, bottom). Together, these results show that local inhibition mediates a stimulus-driven increase in reproducibility of spiking activity.

## Discussion

We recorded the spiking and subthreshold activity of neurons in the high-level auditory regions CM and NCM, in the starling forebrain, in response to spectro-temporally diverse natural songs. We find that song stimuli drive time-varying subthreshold membrane voltage responses that are reliable across trials and stimulus-specific, but non-selective. These results are inconsistent with a model in which neurons with highly selective spiking responses pool over sets of inputs that only drive synaptic activity in service of the selective spiking response. This indicates that, at least these and possibly other high-level sensory neurons are sampling from a much more rich and diverse stimulus space than is evinced by their output spiking. When local inhibition is blocked, spiking variability across trials increases in NCM, suggesting that inhibition plays a key role in governing the stimulus specificity of the time-varying membrane voltage activity generated in a distributed pooling regime.

Understanding how stimulus encoding supports the adaptive interaction between animals and their environment requires studying how natural stimuli are represented in the spiking activity of individual neurons and their populations. The relationship between spiking activity in single neurons and a sensory stimulus depends on both the properties of a potential set of inputs and the pooling operation across those inputs in the post-synaptic cell. In general, the complexity of stimulus selectivity increases along the sensory processing pathway with sparse-firing neurons in ‘higher’ regions selectively spiking in response to progressively more complex features (Sen *et al.*, 2001; Hsu *et al.*, 2004; Meliza *et al.*, 2010; Marshel *et al.*, 2011; DiCarlo *et al.*, 2012). This organization implies a functional hierarchy and presumably confers benefits for the classification and recognition of complex signals (Babadi & Sompolinsky, 2014). To explain higher-order feature selectivity, classic models first developed for primary visual cortex (Hubel & Wiesel, 1962; Movshon *et al.*, 1978) relied on the combination of inputs selective for simpler component features. This kind of feed-forward pooling forms the basis for more contemporary models of object selectivity (Riesenhuber & Poggio, 1999; Rauschecker & Scott, 2009; Lien & Scanziani, 2013), where selective convergence of inputs at each level of a hierarchical network gives rise to narrowed stimulus tuning and more complex receptive field structure. Functionally, these models capture the transformation from dense low-level feature representations to sparse encoding of high-dimensional objects, but the synaptic and network mechanisms underlying this (or other hierarchical models) is not well understood – especially in high-order neurons driven by continuous stimuli comprising diverse sets of features over long periods of time.

Natural signals are ongoing, temporally variable and diverse. We observe reliable spiking activity in response to natural song in regions CM and NCM of the starling cortex, demonstrating that the neurons are clearly auditory. In concert with the temporal variability of the signal, however, spiking is not evenly distributed throughout stimulation. In our own data, the distribution of the stimulus-evoked spike rates averaged over an entire song stimulus is not much different from the mean spontaneous spike rate (stimulus evoked: 1.0 spikes/s, CI = 0.4–5.9 spikes/s; spontaneous: 1.0 spikes/s, CI = 0.3–3.3 spikes/s), demonstrating that the mean spike rate over the song is indeed a poor measure of the cell’s temporally sparse response (Fig. 1). This raises the interesting question of what inputs a cell might receive during periods of the stimulus, when there is no stimulus-specific spiking response. Drawing on the majority of physiology experiments, in which the stimulus has been abstracted from the world and arbitrarily tailored to the response, one might reason that

the inputs revert to a spontaneous state. Indeed, statements in the literature that explicitly or implicitly assume sparsely spiking neurons are not carrying out any computations for the majority of the time are not hard to find (Abeles *et al.*, 1990; Graham & Field, 2007a). But these assumptions have not been tested.

Our results address two potential synaptic pooling regimes, referred to as 'sparse' and 'distributed' pooling, that could underlie sparse, object-selective spiking to natural stimuli. Both pooling regimes can permit selective spiking output, but are distinguished by the selectivity and stimulus-specificity of subthreshold activity throughout long bouts of ongoing natural stimuli. Under sparse pooling, neurons pool inputs that are biased towards the stimulus features that ultimately drive changes in spike rates. For our analyses, this predicts that stimulus-specific activity should be largely restricted, on average, to the response epochs that inform spiking. Conversely, the subthreshold activity that is uninformative of stimulus-specific spiking responses – namely the subthreshold activity not designated as response epochs – should not be stimulus-specific. Our results do not support these predictions. Instead, we find that subthreshold activity is remarkably reliable and stimulus-specific throughout every song presented, regardless of whether there is a stimulus-driven change in the probability of spiking (Figs 1 and 2). This nearly continuous, stimulus-specific control over the subthreshold response rules out a sparse pooling regime.

Instead, the pattern of results favors a distributed pooling regime, in which neurons with selective spiking responses pool synaptic inputs that collectively produce a non-selective, but stimulus-specific, subthreshold response. Importantly, our results cannot address whether the individual pre-synaptic neurons themselves are selectively or non-selectively tuned, but two different scenarios seem most plausible to account for the observed ongoing stimulus-driven reliability and temporal specificity. One possibility is that individual CM and NCM neurons pool inputs from neurons tuned to low-dimensional features that have a relatively high probability of occurring at many points throughout song. These inputs could come from Field L, where some neurons are well-described as frequency-tuned. Their spiking responses are correspondingly non-selective throughout vocalizations and could provide continuous drive to a post-synaptic neuron because power in a given frequency channel is distributed broadly throughout the song. This scenario implies that the populations over which a given neuron is pooling have low population sparseness. Although sparseness appears to be a common feature of sensory encoding (Vinje & Gallant, 2000, 2002; Olshausen & Field, 2004; Graham & Field, 2007b; Hromádka *et al.*, 2008), it may not be the only representational scheme in place (Sakata & Harris, 2009; Tolhurst *et al.*, 2009). A second possibility is that individual NCM and CM neurons pool inputs from neurons tuned to high-dimensional features that occur with relatively low probability throughout the song. In this case inputs might come from other neurons within CM, NCM and other auditory regions that exhibit high population sparseness and whose neurons have high lifetime sparseness. Dense recurrence among (or within) these regions could allow for pooling over much larger numbers of these selective inputs and provide continuous drive to a post-synaptic neuron. In both scenarios, sparse post-synaptic spiking could emerge through the covariance of a specific set of inputs regardless of their tuning. Differentiating between these two possibilities requires a better understanding of the diversity of features to which the inputs are tuned and the combinatorial operations that govern their integration. In either case, our data are consistent with a model in which sparse spiking responses emerge through distributed pooling, which places demands on post-synaptic computational mechanisms to generate

sparse spiking responses. These mechanisms, and their relationship to non-linearities imposed by the spiking threshold, will be important to investigate in future work.

Maintaining synaptic inputs for features that do not co-vary directly with selective spiking output seems somewhat counterintuitive, but it may confer computational advantages for behavior. We know that responses to elements of natural auditory signals are highly dependent on the information conveyed about behavior by particular acoustic material, so the stimulus–response relationship is heavily modulated by behavioral experience (Blake *et al.*, 2002, 2006; Gentner & Margoliash, 2003; Thompson *et al.*, 2010, 2013; Jeanne *et al.*, 2011, 2013; Meliza & Margoliash, 2012). Distributed pooling may confer individual neurons with flexible encoding across a wide diversity of stimulus features, allowing behavioral feedback to shape the pluripotency of the same inputs through synaptic plasticity, adaptation or other mechanisms (Kozlov & Gentner, 2014). Thus, a given neuron may 'represent' multiple objects based on the dynamic functional organization of the system (Kozlov *et al.*, 2013).

The subthreshold activity of all neurons in our sample is consistent with a distributed synaptic pooling architecture. This, in turn, implies that most (if not all) the neurons in these regions are highly interconnected, potentially causing correlations that could lead to deleterious effects on the encoding/decoding of spiking responses across cells (Cohen & Maunsell, 2009; Cohen & Kohn, 2011). Theoretical work has shown, however, that recurrent connectivity in cortical networks can actually decorrelate the activity patterns of neurons with shared presynaptic input (Helias *et al.*, 2014), specifically through inhibitory feedback (Tetzlaff *et al.*, 2012; Bernacchia & Wang, 2013). Correlated firing can also modulate the gain of postsynaptic cells (Salinas & Sejnowski, 2000). We find a major role for inhibition in shaping both the trial-to-trial reproducibility of the post-synaptic response to ongoing natural stimuli, and the magnitude of spike rates – effects that inhibition could manifest by altering the correlation structure of the network. The gain of responses could also be modulated directly by feed-forward inhibition (Mejias *et al.*, 2014). Input–output mappings even in single neurons are not static over time or behavioral conditions (Kozlov & Gentner, 2014), and our results demonstrate that inhibition is poised to provide flexible control over these response characteristics under natural conditions. To understand how inhibition is modulating the sensitivity of the input–output function in neurons of the Starling auditory cortex it will be useful to develop more precise ways to isolate and manipulate different sources of inhibition in the network.

The results of the current study expand our understanding of the synaptic and network mechanisms that underlie hierarchical selective representations of complex natural communication signals in the auditory system. The implications of specific types of synaptic convergence on the computations performed by object-selective neurons in sensory cortex have not been well established and they constrain models for the sparse, selective encoding of natural stimuli. Demands on synaptic plasticity to tune the output of individual neurons from broadly selective synaptic input could provide a computational advantage for increasing the amount of information a sensory signal conveys about behavior across different contexts, improving both the flexibility and the efficiency of stimulus encoding. Network organization and synaptic integration both shape the input–output relationship between the stimulus and the neural response in single cells. Knowing more about these operations in the context of naturally occurring stimuli is a critical step to improving on current models of how stimulus selectivity arises in neural networks.

## Acknowledgements

We thank the members of the Gentner laboratory for comments on the manuscript. The research was funded by grant R01DC008358 from the National Institutes of Health and the Kavli Institute for Brain and Mind to T.Q.G.; and by the Institute for Neural Computation's NIMH Training Program in Cognitive Neuroscience at UCSD to K.E.P. We have no conflicts of interest to declare.

## Abbreviations

CM, caudal mesopallium; GABA, gamma-aminobutyric acid; NCM, caudo-medial nidopallium.

## References

- Ables, M., Vaadia, E. & Bergman, H. (1990) Firing patterns of single units in the prefrontal cortex and neural network models. *Network*, **1**, 13–25.
- Adrian, E.D. (1926) The impulses produced by sensory nerve-endings: Part 4. Impulses from pain receptors. *J. Physiol.*, **62**, 33–51.
- Ayaz, A. & Chance, F.S. (2009) Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J. Neurophysiol.*, **101**, 958–968.
- Babadi, B. & Sompolinsky, H. (2014) Sparseness and expansion in sensory representations. *Neuron*, **83**, 1213–1226.
- Bernacchia, A. & Wang, X.J. (2013) Decorrelation by recurrent inhibition in heterogeneous neural circuits. *Neural Comput.*, **25**, 1732–1767.
- Blake, D.T., Strata, F., Churchland, A.K. & Merzenich, M.M. (2002) Neural correlates of instrumental learning in primary auditory cortex. *Proc. Natl. Acad. Sci. USA*, **99**, 10114–10119.
- Blake, D.T., Heiser, M.A., Caywood, M. & Merzenich, M.M. (2006) Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. *Neuron*, **52**, 371–381.
- Churchland, M.M., Yu, B.M., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S. & Shenoy, K.V. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.*, **13**, 369–378.
- Cohen, M.R. & Kohn, A. (2011) Measuring and interpreting neuronal correlations. *Nat. Neurosci.*, **14**, 811–819.
- Cohen, M.R. & Maunsell, J.H. (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.*, **12**, 1594–1600.
- DiCarlo, J.J., Zoccolan, D. & Rust, N.C. (2012) How does the brain solve visual object recognition? *Neuron*, **73**, 415–434.
- Ganguli, S. & Sompolinsky, H. (2012) Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.*, **35**, 485–508.
- Gentner, T.Q. (2008) Temporal scales of auditory objects underlying bird-song vocal recognition. *J. Acoust. Soc. Am.*, **124**, 1350–1359.
- Gentner, T.Q. & Margoliash, D. (2003) Neuronal populations and single cells representing learned auditory objects. *Nature*, **424**, 669–674.
- Graham, D.J. & Field, D.J. (2007a) Sparse coding in the neocortex. In Kaas, J.H. (Ed.), *Evolution of Nervous Systems*, vol 3. Academic Press, Oxford, pp. 181–187.
- Graham, D.J. & Field, D.J. (2007b) Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spatial Vision*, **21**, 149–164.
- Heliás, M., Tetzlaff, T. & Diesmann, M. (2014) The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.*, **10**, e1003428.
- Hromádka, T., Deweese, M.R. & Zador, A.M. (2008) Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.*, **6**, e16.
- Hsu, A., Woolley, S.M., Fremouw, T.E. & Theunissen, F.E. (2004) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *J. Neurosci.*, **24**, 9201–9211.
- Hubel, D.H. & Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, **160**, 106–154.
- Isaacson, J.S. & Scanziani, M. (2011) How inhibition shapes cortical activity. *Neuron*, **72**, 231–243.
- Jeanne, J.M., Thompson, J.V., Sharpee, T.O. & Gentner, T.Q. (2011) Emergence of learned categorical representations within an auditory forebrain circuit. *J. Neurosci.*, **31**, 2595–2606.
- Jeanne, J.M., Sharpee, T.O. & Gentner, T.Q. (2013) Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, **78**, 352–363.
- Keller, G.B. & Hahnloser, R.H. (2009) Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, **457**, 187–190.
- Kozlov, A.S. & Gentner, T.Q. (2014) Central auditory neurons display flexible feature recombination functions. *J. Neurophysiol.*, **111**, 1183–1189.
- Kozlov, A.S., Briggs, A., Gentner, T.Q. & Sharpee, T.O. (2013) *Mosaic receptive fields of central auditory neurons optimized to the natural sound statistics*. Paper presented at the Society for Neuroscience, San Diego.
- Lien, A.D. & Scanziani, M. (2013) Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.*, **16**, 1315–1323.
- Margrie, T.W., Brecht, M. & Sakmann, B. (2002) In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflug. Arch.*, **444**, 491–498.
- Marshel, J.H., Garrett, M.E., Nauhaus, I. & Callaway, E.M. (2011) Functional specialization of seven mouse visual cortical areas. *Neuron*, **72**, 1040–1054.
- Mejias, J.F., Payeur, A., Selin, E., Maler, L. & Longtin, A. (2014) Subtractive, divisive and non-monotonic gain control in feedforward nets linearized by noise and delays. *Front. Comput. Neurosci.*, **8**, 19.
- Meliza, C.D. & Margoliash, D. (2012) Emergence of selectivity and tolerance in the avian auditory cortex. *J. Neurosci.*, **32**, 15158–15168.
- Meliza, C.D., Chi, Z. & Margoliash, D. (2010) Representations of conspecific song by starling secondary forebrain auditory neurons: toward a hierarchical framework. *J. Neurophysiol.*, **103**, 1195–1208.
- Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.*, **283**, 53–77.
- Olshausen, B.A. & Field, D.J. (2004) Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.*, **14**, 481–487.
- Rauschecker, J.P. & Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.*, **12**, 718–724.
- Riesenhuber, M. & Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, **2**, 1019–1025.
- Sakata, S. & Harris, K.D. (2009) Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron*, **64**, 404–418.
- Salinas, E. & Sejnowski, T.J. (2000) Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.*, **20**, 6193–6209.
- Schneider, D.M. & Woolley, S.M. (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron*, **79**, 141–152.
- Sen, K., Theunissen, F.E. & Doupe, A.J. (2001) Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.*, **86**, 1445–1458.
- Tetzlaff, T., Heliás, M., Einevoll, G.T. & Diesmann, M. (2012) Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.*, **8**, e1002596.
- Thompson, J.V., Jeanne, J. & Gentner, T.Q. (2010) *Local inhibition shapes the learned responses to song in NCM*. Paper presented at the Soc. Neurosci Abstracts.
- Thompson, J.V., Jeanne, J.M. & Gentner, T.Q. (2013) Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. *J. Neurophysiol.*, **109**, 721–733.
- Tolhurst, D.J., Smyth, D. & Thompson, I.D. (2009) The sparseness of neuronal responses in ferret primary visual cortex. *J. Neurosci.*, **29**, 2355–2370.
- Vinje, W.E. & Gallant, J.L. (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, **287**, 1273–1276.
- Vinje, W.E. & Gallant, J.L. (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.*, **22**, 2904–2915.
- Wilent, W.B. & Contreras, D. (2005) Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat. Neurosci.*, **8**, 1364–1370.
- Zhu, J.J., Uhlich, D.J. & Lytton, W.W. (1999) Properties of a hyperpolarization-activated cation current in interneurons in the rat lateral geniculate nucleus. *Neuroscience*, **92**, 445–457.