

# Long-Lasting Context Dependence Constrains Neural Encoding Models in Rodent Auditory Cortex

Hiroki Asari and Anthony M. Zador

Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Cold Spring Harbor, New York

Submitted 6 July 2009; accepted in final form 12 August 2009

**Asari H, Zador AM.** Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J Neurophysiol* 102: 2638–2656, 2009. First published August 12, 2009; doi:10.1152/jn.00577.2009. Acoustic processing requires integration over time. We have used in vivo intracellular recording to measure neuronal integration times in anesthetized rats. Using natural sounds and other stimuli, we found that synaptic inputs to auditory cortical neurons showed a rather long context dependence, up to  $\geq 4$  s ( $\tau \sim 1$  s), even though sound-evoked excitatory and inhibitory conductances per se rarely lasted  $\geq 100$  ms. Thalamic neurons showed only a much faster form of adaptation with a decay constant  $\tau < 100$  ms, indicating that the long-lasting form originated from presynaptic mechanisms in the cortex, such as synaptic depression. Restricting knowledge of the stimulus history to only a few hundred milliseconds reduced the predictable response component to about half that of the optimal infinite-history model. Our results demonstrate the importance of long-range temporal effects in auditory cortex and suggest a potential neural substrate for auditory processing that requires integration over timescales of seconds or longer, such as stream segregation.

## INTRODUCTION

One goal of systems neuroscience is to characterize the relationship between input sensory stimuli and output neural responses. Linear models have been widely used in the auditory and visual systems due to their simplicity and interpretability (Eggermont et al. 1983; Escabí and Schreiner 2002; Klein et al. 2000; Theunissen et al. 2001; Wu et al. 2006). Linear spectrotemporal receptive field (STRF) models have been quite successful in describing the input–output function of some stimulus ensembles in auditory cortex (Depireux et al. 2001; Kowalski et al. 1996), but have yielded only poor results for other ensembles, including those consisting of natural stimuli or other complex stimuli (Linden et al. 2003; Machens et al. 2004).

Why has the classical STRF-based approach failed to provide a general model? The straightforward answer is that the actual input–output function is nonlinear. For example, the actual input–output function might include multiplicative interactions between different frequency bands. However, the space of nonlinear functions is large and it is not feasible to fit general high-order models. For instance, if the input spectrogram is naïvely discretized with a (rather coarse) frequency resolution of 0.25 octave over 5 octaves and a (rather coarse) temporal resolution of 10 ms over 200 ms, then the number of parameters for a linear model is:  $N = 5/0.25 \times 200/10 = 400$ ,

whereas it is  $\mathcal{O}(N^2)$  (about 160,000) for a second-order Wiener model and, in general,  $\mathcal{O}(N^n)$  for an  $n$ th-order Wiener model. The success of general “black-box” nonlinear models is thus quickly limited by the “curse of dimensionality,” i.e., the fact that the amount of data required to fit a general model increases exponentially with the order of the model; even general black-box models that are guaranteed to succeed in principle are often data-limited in practice. Although in some cases the difficulties can be circumvented by the judicious choice of nonlinearities (Ahrens et al. 2008; Chichilnisky 2001; Fishbach et al. 2001, 2003; Rust et al. 2005; Schwartz and Simoncelli 2001; Sharpee et al. 2004), it is difficult to know a priori what form the nonlinearities should take.

One way to reduce the number of model parameters is to tailor the model to the observed properties of auditory cortex neurons. The preceding parameter count illustrates that the system’s “memory”—i.e., the dependence of the neuron’s input–output behavior on stimulus history or context—is one of the primary determinants of model complexity; doubling the length of the memory (e.g., from 200 to 400 ms in the above-cited example) doubles the number of input variables (from  $N$  to  $2N$  for fixed temporal resolution). Thus it would be useful to characterize the length of the system’s memory.

Here we provide for the first time a quantification of long-lasting stimulus context effects in determining the stimulus–response properties of single neurons in the primary auditory cortex. We used in vivo whole cell patch-clamp recordings in anesthetized rats to examine subthreshold responses in a paradigm in which a given probe stimulus was preceded by different conditioning stimuli. These conditioning stimuli provided a temporal context. Both probe and conditioning stimuli were drawn from natural and synthetic sound ensembles with rich temporal and spectral structure. We found that context dependence could last for a rather long time—sometimes as long as  $\geq 4$  s. The long-lasting effects described are elicited by a much broader range of stimuli than those described in an animal model of stimulus specific adaptation (Pienkowski and Eggermont 2009; Ulanovsky et al. 2003, 2004), suggesting that they represent a much more general phenomenon. Consistent with previous results (Creutzfeldt et al. 1980; Miller et al. 2002; Ulanovsky et al. 2003, 2004; Wehr and Zador 2005), this long-lasting context dependence originated in cortex at the level of synaptic inputs and was not seen in thalamus. Extending the memory of linear models did not improve their performance, indicating that these long-lasting effects of context were nonlinear. The slow stimulus adaptation we report may play a role in stream segregation and other forms of auditory processing that require integration over seconds.

Address for reprint requests and other correspondence: A. M. Zador, Cold Spring Harbor Laboratory, Watson School of Biological Sciences, One Bungtown Road, Cold Spring Harbor, NY 11724 (E-mail: zador@cshl.edu).

## METHODS

We performed all data analysis in MATLAB (The MathWorks, Natick, MA).

*Surgery*

Long-Evans rats (20–28 days old) were anesthetized (30 mg/kg ketamine and 0.24 mg/kg medetomidine) in strict accordance with the National Institutes of Health guidelines as approved by the Cold Spring Harbor Laboratory Animal Care and Use Committee. After the animal was deeply anesthetized, it was placed in a custom nasoorbital restraint, which left the ears free and clear. A cisternal drain was made and a small craniotomy and durotomy were performed above the left primary auditory cortex (area A1). The cortex was covered with physiological buffer (in mM: NaCl, 127; Na<sub>2</sub>CO<sub>3</sub>, 25; NaH<sub>2</sub>PO<sub>4</sub>, 1.25; KCl, 2.5; MgCl<sub>2</sub>, 1; and glucose, 25) mixed with 1.5% agarose. Temperature was monitored rectally and maintained at 37°C using a feedback-controlled blanket. Depth of anesthesia was monitored throughout the experiment and supplemental anesthesia was provided when required.

*Electrophysiology*

Whole cell and cell-attached recordings were obtained *in vivo* using standard blind patch-clamp recording techniques (see, e.g., Machens et al. 2004; Wehr and Zador 2003, 2005). Electrodes were pulled from filamented, thin-walled, borosilicate glass (outer diameter, 1.5 mm; inner diameter, 1.17 mm; World Precision Instruments, Sarasota, FL) on a vertical two-stage puller (Narishige, East Meadow, NY). Internal solution for current-clamp recordings contained (in mM): KCl, 20; K-gluconate, 100; HEPES, 10; MgCl<sub>2</sub>, 2; CaCl<sub>2</sub>, 0.05; Mg-ATP, 4; Na<sub>2</sub>-GTP, 0.3; Na<sub>2</sub>-phosphocreatine, 10; and about 2.5 micro-emerald (dextran-conjugated fluorescent dye; Invitrogen); pH 7.3; diluted to 275 mOsm. For voltage-clamp recordings, we used the following internal solution (in mM) to pharmacologically block action potentials: K-gluconate, 140; HEPES, 10; MgCl<sub>2</sub>, 2; CaCl<sub>2</sub>, 0.05; Mg-ATP, 4; Na<sub>2</sub>-GTP, 0.4; Na<sub>2</sub>-phosphocreatine, 10; BAPTA, 10; and QX-314, 5; pH 7.25; diluted to 290 mOsm, producing a calculated reversal potential of –85 mV for both K<sup>+</sup> and Cl<sup>–</sup> conductances. Resistance to bath was 3.5–5.0 MΩ before seal formation. We used a custom data acquisition system written in MATLAB and sampled membrane

potential at 10 kHz using an amplifier Axopatch 200B (Molecular Devices, Palo Alto, CA) in current- or voltage-clamp mode with no on-line series resistance compensation. Mean series resistance was 30.0 ± 12.3 MΩ (mean ± SD; 16 cells) for cell-attached recordings and 68.8 ± 16.7 MΩ (mean ± SD; 189 cells) for current-clamp recordings. For voltage-clamp recordings, mean input and series resistances were 142.8 ± 68.7 and 74.8 ± 24.3 MΩ (mean ± SD; 42 cells), respectively. Holding potentials were stepped (using a 1-s ramp) to a pseudorandom sequence of three values. At each potential, after a 1-s equilibration period, ten 10-mV square pulses (30-ms duration) were delivered at 12.5 Hz to monitor input and series resistances, followed by acoustic stimuli.

Whole cell recordings were made from primary auditory cortex (area A1) as determined by the tonotopic gradient and by the frequency–amplitude tuning properties of cells and local field potentials. We recorded from almost all subpial depths (range: 76–936 μm, as determined from micromanipulator travel). Thirteen cells were recovered histologically, which were verified to be pyramidal cells (e.g., Fig. 2C). All together, we recorded from 194 cells in 139 animals in current-clamp mode, of which 123 cells met our criterion for the analysis (see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL). Of these, 39 cells were examined with natural sound ensembles, 14 cells with ensembles of temporally orthogonal ripple combinations, 6 cells with ensembles of dynamic moving ripples, 39 cells with ensembles of modulated harmonic tones, and 27 cells with ensembles of modulated colored noise ensembles (Table 1). In voltage-clamp mode we recorded from 42 cells in 17 animals, of which we analyzed 14 cells that were probed with natural sound ensembles (see CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL).

Cell-attached recordings in the thalamus (3.62–4.35 mm deep from the surface of area A1) were obtained from the ventral division of the medial geniculate body (MGB) as determined by short latency (8.7 ± 3.1 ms; mean ± SD) and the “V-shaped” frequency response areas. In total we recorded 16 cells in five animals and examined all with natural sound ensembles. Of these, 14 cells met our criterion for the analysis (see CONTEXT DEPENDENCE AT SUPRATHRESHOLD LEVEL).

*Stimulus design*

During the recordings, we presented sequences of various stimulus combinations in a randomly interleaved manner (see *Stimuli* for

TABLE 1. Summary of recording data

Probe	Sound Properties Varied Among Conditioning Stimuli					
	All	Amplitude	Frequency	AM	FM	Higher-Order
A1						
Natural sound (VC)	29 (14)					
Natural sound (CC)	305 (39)	23 (23)	25 (25)	25 (25)	25 (25)	63 (27)
TORC		39 (9)	40 (9)			
DMR		8 (2)	20 (5)			
MHT		25 (20)	25 (21)	71 (27)	57 (25)	
MCN						74 (27)
MGB						
Natural sound	93 (14)					

Shown is the number of probe stimuli tested in at least two different conditioning stimuli repeated over at least 2 or 4 trials in A1 for voltage- or current-clamp recordings, respectively, and at least 10 trials in MGB. Each row represents the probe stimulus type, whereas each column indicates the stimulus properties varied among the conditioning stimuli. The corresponding number of recorded cells in A1 or MGB is shown in parentheses. A given cell could be tested with more than one probe stimulus, each of which could in turn be tested with more than one type of conditioning stimulus ensemble (see also *Stimulus design* in METHODS and Fig. 1). *Natural sounds* would differ in *all* possible sound properties, allowing us to examine overall context-dependence effects (Figs. 2–6 and 9), whereas synthetic sounds (*TORC*, *DMR*, *MHT*, and *MCN*) were used to examine the effects caused by the changes in each of the following sound properties among conditioning stimuli (Fig. 8); *amplitude* over the maximum range of 60-dB attenuation, *frequency* each with the maximum shift of 4 octaves (e.g., Fig. 7), *AM* and *FM*, with the maximum difference of 20 Hz (rate) and threefold (depth), and *higher-order* properties by comparing natural sounds and corresponding modulated colored noise. VC, voltage-clamp; CC, current-clamp; TORC, temporally orthogonal ripple combination; DMR, dynamic moving ripple; MHT, modulated harmonic tone; MCN, modulated colored noise; AM, amplitude-modulation; FM, frequency-modulation; A1, primary auditory cortex; MGB, medial geniculate body.

details of stimulus fragments). To maximize the yield in finite recording length (typically  $\sim 20$  and  $\sim 40$  min for whole cell and cell-attached recordings, respectively), we generated fixed  $N$  sequences from  $N$  stimulus fragments ( $S_i$  for  $i = 1, \dots, N$ ) that allow for examining the responses to all stimulus pairs ( $S_i S_j$ ; conditioning stimulus  $S_i$ , probe stimulus  $S_j$ ) and to each stimulus following a “silent” period, i.e., an intersequence interval. Formally, stimulus sequences we presented (given  $N$  stimulus fragments and a “silent” period) follow a *cyclic code* over a finite field  $\mathbb{F}_{N+1}$  of block length two (van Lint 1992). For  $N = 4$ , for example, we would present the following four stimulus sequences:  $S_1 S_3 S_2 S_4 S_4$ ,  $S_3 S_4 S_1 S_1 S_2$ ,  $S_4 S_2 S_3 S_3 S_1$ , and  $S_2 S_2 S_1 S_4 S_3$  (Fig. 1A). Interstimulus intervals and intersequence intervals were 0 and about 6 s, respectively.

In this circular stimulus design, any stimulus fragment—including “silence”—can thus be considered as both a probe stimulus (for the previous context) and a conditioning stimulus (for the following probe). Care should be taken, however, for the analysis (Fig. 1A). First, even though a subset of cells showed significant context-dependence effects soon after the onset of intersequence intervals (Supplemental Fig. S1A),<sup>1</sup> we did not include the analysis results on such “silence probe” periods in Fig. 3 because the estimated response power attributable to any “context stimulus” alone faded away within  $< 1$  s after the stimulus termination (Supplemental Fig. S1B). Second, although any stimulus can be considered as a conditioning stimulus for analyzing overall context-dependence effects (Figs. 2–6 and 9), only an appropriate set of synthetic stimulus fragments can be used as conditioning stimuli for analyzing the contribution of individual

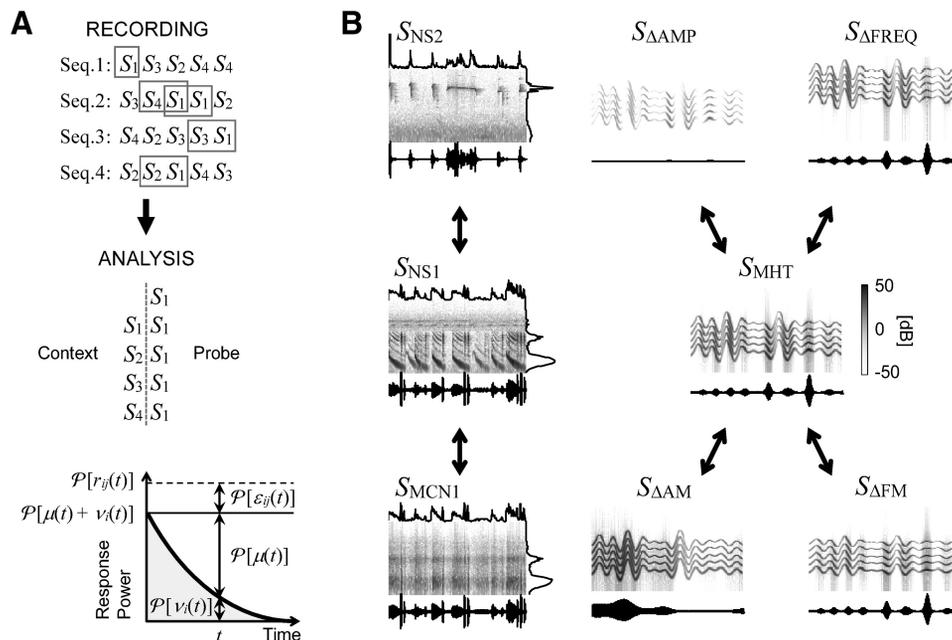
acoustic properties (Figs. 7 and 8). For example, to assess the frequency change effects with three synthetic frequency variants ( $S_1$ ,  $S_2$ , and  $S_3$ ) and one natural sound fragment ( $S_4$ ), the cyclic stimulus design in Fig. 1A for instance can be used for the recordings, but the analysis should be conducted by using only those frequency variants ( $S_i$  for  $i = 1, \dots, 3$ ) as conditioning stimuli for each probe fragment ( $S_i$  for  $i = 1, \dots, 4$ ; see also Fig. 7 and Supplemental Fig. S9).

### Stimuli

All stimuli were delivered at 97.656 or 200 kHz using a TDT System 3 with an ED1 electrostatic speaker (Tucker-Davis Technologies, Alachua, FL) in free-field configuration (speaker located  $\sim 8$  cm lateral to, and facing, the contralateral ear) in a double-walled sound booth (Industrial Acoustics, Bronx, NY). The speaker had a maximum intensity (at 10-V command voltage) of 92-dB sound pressure level (SPL) and its frequency response was flat from 1 to 22 kHz to within SD of 3.7 dB. Sound levels were measured with a type 7012 1/2-in. ACO Pacific microphone (ACO Pacific, Belmont, CA) positioned where the contralateral ear would be (but with the absence of animal).

**NATURAL SOUNDS.** Natural sound ensembles were used to assess the overall context-dependent effects (Figs. 2–6 and 9). All natural sound fragments were taken from commercially available audio compact discs, originally sampled at 44.1 kHz and resampled at 97.656 or 200 kHz for stimulus presentation: *The Diversity of Animal Sounds and Sounds of Neotropical Rainforest Mammals* (Cornell Laboratory of Ornithology, Ithaca, NY). The majority of the sound sections lasted

<sup>1</sup> The online version of this article contains supplemental data.



**FIG. 1.** Experimental design and auditory stimuli. **A:** experimental design for analyzing context dependence. During the *recording*, we presented well-designed  $N$  sequences of a given set of  $N$  sound fragments with no interstimulus interval in a randomly interleaved manner (shown is an example for  $N = 4$ ; intersequence interval,  $\sim 6$  s). For the *analysis*, we aligned the recording data to examine the variability in the responses to a given sound fragment (*probe*;  $S_1$  in this example) due to the presence of different preceding stimuli (*context*; “silence,”  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_4$ ). The choice of conditioning stimuli depends on the goal of the analysis (for details, see *Stimulus design* in METHODS). Here we assumed that the response power (broken line;  $\mathcal{P}[r_{ij}(t)]$  from Eq. 11) to a probe stimulus at time  $t$  from probe onset can be divided into noise power ( $\mathcal{P}[\varepsilon_{ij}(t)]$ ) and stimulus-related power (thin black;  $\mathcal{P}[\mu(t) + v_i(t)]$  in Eq. 17) that can be further decomposed into a context-independent fraction ( $\mathcal{P}[\mu(t)]$  in Eq. 18) and a context-dependent fraction (gray;  $\mathcal{P}[v_i(t)]$  in Eq. 19). For details, see *Analysis* in METHODS. **B:** natural sounds and synthetic sounds. Natural sound fragments ( $S_{NS1}$  and  $S_{NS2}$ ; 4.11 s long; sound pressure waveforms, spectrograms, and temporal and spectral marginal distributions) differ substantially in their spectrotemporal components, which causes a large and long context dependence in A1 when they are used as conditioning stimuli (Figs. 2 and 3). On the other hand, the temporal and spectral patterns in the marginal distributions between modulated colored noise  $S_{MCN1}$  and the corresponding natural sound  $S_{NS1}$  are nearly identical, resulting in a small and short context dependence (Fig. 8). Synthetic sounds such as modulated harmonic tones (sound pressure waveform and corresponding spectrograms; 1 s long) can be used to assess the effects of the changes in sound properties in more detail. Compared with  $S_{MHT}$ , for example,  $S_{\Delta AMP}$  has 30 dB less power,  $S_{\Delta FREQ}$  has the frequency components up-shifted by 1.5 octaves,  $S_{\Delta AM}$  has slower AM rates by 4 Hz on average, and  $S_{\Delta FM}$  has half the SD for the frequency-modulated (FM) depth. For details, see *SYNTHETIC SOUNDS* in METHODS.

for 3.5–6.5 s, but some were shorter (2–3 s), to examine as many stimulus combinations as possible (see *Stimulus design* and Fig. 1A). The sound segments were chosen from original sound tracks to have minimum “silent” periods (especially at the onset and termination) and a 5-ms cosine-squared ramp was applied at the onset and termination to ensure a smooth connection between the segments, even with no interstimulus interval. The peak amplitude of each segment was normalized to the  $\pm 10$ -V range of the speaker driver. The natural sound stimuli consisted of 46 different sounds in total—covered almost all frequencies from 0 to 22 kHz and ranged from narrow- to broadband stimuli—although only a subset of stimuli was tested on any particular cell. We typically presented combinations of  $N \sim 7$  different stimuli on each cell for current-clamp and cell-attached recordings (Figs. 2–4 and 7–9). For voltage-clamp recordings, we typically examined a few probe stimuli on each cell, where each probe stimulus was preceded by three to six different conditioning stimuli (including “silence” stimulus; Figs. 5 and 6).

**SYNTHETIC SOUNDS.** Synthetic sound ensembles—with or without one additional natural sound fragment for a probe—were used to examine the effects of the changes in each of the following acoustic properties (Figs. 7 and 8): intensity (amplitude), frequency, amplitude-modulation (AM), frequency-modulation (FM), and higher-order spectrotemporal acoustic features. We used temporally orthogonal ripple combinations (TORCs; for details, see, e.g., Klein et al. 2000) and dynamic moving ripples (DMRs; for details, see, e.g., Escabí and Schreiner 2002) to examine how changes in amplitudes and frequencies in conditioning stimuli contribute to the context-dependence effects (over the maximum range of 40 dB and 4 octaves, respectively; *Eqs. 1–3*). Modulated harmonic tones—generated by combining a time-varying envelope and a harmonic series of time-varying frequencies (Fig. 1B; *Eqs. 4 and 5*)—were used to assess not only the effects of AM and FM changes (with the maximum difference of 20 Hz and threefold in modulation rate and depth, respectively) but also the changes in amplitudes and frequencies (over the maximum range of 60 dB and 2 octaves, respectively). To test the effects of the changes in higher-order sound properties such as complex interactions between spectrotemporal constituents, we used modulated colored noise that has asymptotically the same temporal and spectral patterns in the marginal distribution of the spectrogram as those of a target stimulus (*Eqs. 6–9*). All synthetic sounds were sampled at 97.656 or 200 kHz and lasted for 4.0–5.5 s (Fig. 1B and Table 1).

In this study, no cell was tested with all the sound properties in conditioning stimuli due to a limited recording length. Strictly speaking, then, we cannot directly compare the context-dependence effects caused by the changes in different acoustic properties. However, the comparisons we made at the population level (Fig. 8) would be reasonable because the acoustic properties were varied across almost the entire range that A1 neurons can follow faithfully (e.g., sound trains or temporal modulations up to tens of Hertz; Creutzfeldt et al. 1980; Joris et al. 2004).

**Temporally orthogonal ripple combinations (TORCs).** The following equation was used to generate ripples and their combinations (Klein et al. 2000)

$$y(t) = \sum_i y_E(t, x_i) \cdot y_C(t, f_i) \quad (1)$$

where the envelope  $y_E(t, x_i)$  and carriers  $y_C(t, f_i)$  are respectively given by

$$20 \log_{10} [y_E(t, x_i)] = a_0 + \sum_{j,k} \frac{a_{jk}}{2} \cos [2\pi(\psi_j(t) + \Omega_k(t)x_i) + \Phi_{jk}] \quad (2)$$

$$y_C(t, f_i) = \sin [2\pi f_i t + \varphi_i] \quad (3)$$

Note that  $a_{jk}$  ( $>0$ ) is a sinusoidal modulation depth around the mean  $a_0$  in dB,  $\omega_j(t) = \partial\psi_j/\partial t$  in Hz and  $\Omega_k(t)$  in cycles/octave are temporal and spectral ripple modulations, respectively,  $x_i = \log_2 [f_i/f_0]$  in octaves is a logarithmic frequency axis relative to  $f_0$  in Hz and  $\Phi_{jk}$  and  $\varphi_i$  are random initial phases.

For generating TORC-based stimuli, envelopes of seven “short” TORCs were first generated, each consisting of six ripples with temporal modulation:  $\omega_j = 4j$  Hz (for  $j = 1, \dots, 6$ ), and each having a fixed ( $k = 1$ ) spectral modulation:  $\Omega = -1.5, -0.9, -0.3, 0, 0.6, 1.2$ , and  $1.8$  cycles/octave, respectively. All TORCs had rise and fall times of 5 ms, modulation depth of  $\leq 30$  dB (with  $a_{jk} = 30/6$ ), and lasted for 250 ms. Such short TORC envelopes were then randomly adjoined to generate a “default envelope” that lasted for 4–5.5 s.

To examine the effects of the changes in sound intensities in conditioning stimuli, we applied the default envelope to the carrier frequencies over a bandwidth of 5 octaves (0.88–28.16 kHz in steps of 1/128 octaves), scaled the peak amplitude to the speaker driver range, and then varied the amplitudes over the maximum range of 40-dB attenuation. To examine the effects of frequency changes, we generated a default envelope over the bandwidth of 2 or 4 octaves and chose carrier sinusoids within the range of 6 octaves (0.625–40 kHz in steps of 1/128 octaves, random phase at *time 0*) so that the signals had the same envelope with shifted bandwidth (e.g., 0.625–10 kHz, 2.5–40 kHz, and so on). We then normalized the peak amplitude of sound fragments with respect to their total signal powers ( $\int |y(t)|^2 dt$ ) and uniformly scaled to fit them all within the speaker driver range.

**Dynamic moving ripples (DMRs).** To generate the DMR envelopes (*Eq. 2*;  $j = k = 1$ ), spectral modulations  $\Omega(t)$  were sampled at 6 Hz from a uniform distribution in interval  $\pm 1.5$  cycles/octave and temporal modulations  $\omega(t)$  were sampled at 3 Hz from a uniform distribution ranging between +25 and –25 Hz, both of which were then up-sampled to 97.656 or 200 kHz using a cubic interpolation procedure (interp1 function with the “cubic” option in MATLAB; see also Escabí and Schreiner 2002).

Carrier frequencies were chosen as in the TORC stimuli and applied to the envelope as in *Eqs. 1–3*. All DMR signals had rise and fall times of 5 ms and modulation depth of  $\leq 30$  dB. The peak amplitude was scaled in the same way as the TORC fragments (see previous subsection). Figure 7A and Supplemental Fig. S9 show example spectrograms of DMR frequency variants.

**Modulated harmonic tones (MHTs).** We used the following equation to generate modulated harmonic tones

$$y_{\text{MHT}}(t) = A(t) \sum_{i=0}^{M-1} \cos [2^{i/m} \phi(t) + \varphi_i] \quad (4)$$

where  $A(t)$  is the envelope and  $\varphi_i$  and  $\phi(t)$  are the initial and time-varying phases, respectively. In this study,  $M = 5$  tones were combined with the density:  $m = 0.5$  or 1 tone/octave. The derivative of the phase  $\phi(t)$  with respect to time gives the instantaneous frequency  $f(t)$

$$\frac{\partial \phi(t)}{\partial t} = 2\pi f(t) \quad (5)$$

Normal distributions sampled at 48 Hz were used to generate the envelope  $A(t)$  and the instantaneous frequency  $f(t)$ . The mean  $A(t)$  ranged between 40 and 65 dB (SD; from 5 to 15 dB), whereas the mean  $f(t)$  ranged over 3 octaves (from 0.375 to 3 kHz) with the SD from 1 to 1/3 octaves. We then up-sampled  $A(t)$  and  $f(t)$  to 97.656 or 200 kHz using a cubic interpolation procedure and used *Eqs. 4 and 5* to generate the signal  $y_{\text{MHT}}(t)$  with random initial phase  $\varphi_i$ .

To examine the effects of amplitude changes, we generated a signal for fixed  $A(t)$  and  $f(t)$ , normalized its peak amplitude within the

speaker driver range, and varied the amplitudes over the maximum range of 60-dB attenuations (see, e.g.,  $S_{\text{MHT}}$  vs.  $S_{\text{DAMP}}$  in Fig. 1B). To examine the effects of frequency changes, we generated signals for fixed  $A(t)$  but with shifted  $f(t)$  by up to  $\pm 2$  octaves, normalized their total signal powers, and uniformly scaled the signals to fit them all within the speaker driver range (see, e.g.,  $S_{\text{MHT}}$  versus  $S_{\text{DFREQ}}$  in Fig. 1B). To examine the effects of the changes in AM or FM, we used fixed mean  $A(t)$  and  $f(t)$ . Before the up-sampling procedures, however, either  $A(t)$  or  $f(t)$  was scaled to vary the modulation depth by up to threefold and/or band-pass filtered (bandwidth: 4 Hz) to limit the modulation rates from [0–4] to [20–24] Hz (see, e.g.,  $S_{\text{MHT}}$  vs.  $S_{\text{DAM}}$  and  $S_{\text{DFM}}$  in Fig. 1B, respectively). The synthetic signals were normalized with respect to their total power and uniformly scaled to fit them all within the speaker driver range.

**Modulated colored noise (MCN).** Starting from white noise  $x_0(t)$ , we used the following iterative procedures to produce modulated colored noise  $y_{\text{MCN}}(t)$  that has asymptotically the same temporal and spectral modulation patterns as that of a target natural sound  $y_{\text{NS}}(t)$ . First, we computed the analytic signal of  $y_{\text{NS}}(t)$  by using the Hilbert transform  $\mathcal{H}[\cdot]$  and decomposed it into the envelope  $A_{\text{NS}}(t)$  and the phase  $\phi_{\text{NS}}(t)$

$$y_{\text{NS}}(t) + \mathcal{H}[y_{\text{NS}}(t)] = A_{\text{NS}}(t) \exp[j\phi_{\text{NS}}(t)] \quad (6)$$

where  $j^2 = -1$ . Second, using the Fourier transform  $\mathcal{F}[\cdot]$ , we filtered the signal from the  $(i - 1)$ th iteration  $x_{i-1}(t)$  to have the same power spectrum as  $y_{\text{NS}}(t)$

$$\tilde{z}_i(\omega_k) = \tilde{x}_{i-1}(\omega_k) \cdot \frac{|\tilde{y}_{\text{NS}}(\omega_k)|}{|\tilde{x}_{i-1}(\omega_k)|} \quad (7)$$

where  $\tilde{y}(\omega) = \mathcal{F}[y(t)]$  denotes the signal  $y$  in the Fourier domain. Third, we computed the analytic signal of  $z_i(t) = \mathcal{F}^{-1}[\tilde{z}_i(\omega)]$  as in Eq. 6

$$z_i(t) + \mathcal{H}[z_i(t)] = B_i(t) \exp[j\psi_i(t)] \quad (8)$$

where  $B_i(t)$  and  $\psi_i(t)$  are the envelope and the phase, respectively. Finally, we generated the signal for the  $i$ th update as

$$x_i(t) = A_{\text{NS}}(t) \cos[\psi_i(t)] \quad (9)$$

That is,  $x_i(t)$  is a colored noise with the envelope of the target  $y_{\text{NS}}(t)$ . In this study, we updated the synthetic signal 1,000 times to generate modulated colored noise:  $y_{\text{MCN}}(t) = x_{1,000}(t)$ . The signals  $y_{\text{NS}}(t)$  and  $y_{\text{MCN}}(t)$  were normalized with respect to their total power and then uniformly scaled to fit them all within the speaker driver range (see, e.g.,  $S_{\text{NS1}}$  and  $S_{\text{MCN1}}$  in Fig. 1B, respectively).

## Analysis

All data analysis was done in the discrete time domain (of resolution:  $\Delta t = 0.1$  ms = 1/sampling rate), but in the following text we omit the indices for time bins for brevity.

For auditory cortical voltage responses, as a preprocessing we applied a median filter (10-ms window) to clip spikes from the raw data and centered the subthreshold responses to have zero mean [i.e.,  $r(t) - \langle r(t) \rangle$ ], instead of subtracting the resting potential;  $\langle \cdot \rangle$ , indicates the average over time  $t$ . Note that this filter operation preserves the subthreshold voltage fluctuations (e.g., compare Fig. 2 and Supplemental Fig. S2). No preprocessing was applied to measured current responses (CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL). Because of the low firing rates in A1 (spontaneous,  $0.47 \pm 0.61$  Hz; evoked,  $0.57 \pm 0.77$  Hz; mean  $\pm$  SD, 194 cells; see, e.g., Supplemental Figs. S2 and S9), we did not perform any further analysis at the spike level.

For auditory thalamic responses, spikes recorded in cell-attached mode were extracted from raw voltage traces by applying a high-pass

filter and thresholding. Spike times were then assigned to the peaks of suprathreshold segments. Sufficiently high firing rates in MGB allowed us to analyze the context dependence at the spike level (spontaneous,  $0.78 \pm 1.25$  Hz; evoked,  $11.4 \pm 16.9$  Hz; mean  $\pm$  SD, 16 cells; see, e.g., Supplemental Fig. S10).

**CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL.** For those current-clamp recordings that we could present at least four repeats of any probe stimulus tested with at least two conditioning stimuli, temporal context dependence—i.e., the response variability to a probe stimulus due to the presence of different conditioning, preceding stimuli—was examined in two ways: 1) significance measure in the statistics sense and 2) fractional power measure in the response dynamics. The relevant timescale was then measured by fitting (a sum of) exponential processes to the population data (EXPONENTIAL CURVE FIT).

**Significance measure.** For each sampled time point  $t$  ( $\geq 0$ ) on a probe stimulus (with  $t = 0$  indicating the transition from conditioning to probe stimuli), we performed a one-way nonparametric ANOVA (Kruskal–Wallis test; Kruskal and Wallis 1952) for equal medians among the subthreshold cortical responses  $r_{ij}(t)$  over trials  $j = 1, \dots, m$  in all the conditioning stimuli  $i = 1, \dots, n$  (i.e., the stimulus contexts we examined; see also Eq. 11). Briefly, we first ranked the data  $r_{ij}(t)$  into integers  $\mathbf{r}_{ij}(t) = 1, \dots, mn$  for each time point  $t$  on a probe stimulus. The test statistic  $KW(t)$  is then given as

$$KW = (mn - 1) \frac{\sum_i m (\langle \mathbf{r}_{ij} \rangle_j - \langle \langle \mathbf{r}_{ij} \rangle_i \rangle)^2}{\sum_i \sum_j (\mathbf{r}_{ij} - \langle \langle \mathbf{r}_{ij} \rangle_i \rangle)^2} = \frac{3 \langle (2 \langle \mathbf{r}_{ij} \rangle_j - mn - 1)^2 \rangle_i}{mn + 1} \quad (10)$$

where  $\langle \cdot \rangle$  indicates the average over trials with subscript  $j$  and over contexts with subscript  $i$  and the probability distribution of  $KW$  can be approximated as a chi-square distribution with  $n - 1$  degrees of freedom. The time index ( $t$ ) is ignored here for brevity.

Our criterion for the significance level was  $P < 0.01$  for  $\geq 5$  ms (i.e.,  $\geq 50$  consecutive time points to avoid false positives due to multiple comparisons over time). In the population data analysis (Figs. 3A and 8A), this significance measure was used to compute the proportion—or probability—of observing significant context dependence at a given moment after probe onset. The noise floor—or, the level of false positive—was determined by resampling methods, in which the trials were randomly shuffled to lose the information on the contexts, followed by the same significance test described earlier. For each probe, we repeated this procedure 1,000 times and took the average over the population to identify the chance level of declaring “significance” in this analysis.

In this study, we did not perform any post hoc analysis partly because of data limitation and partly because our goal of this analysis was to detect whether neurons showed context dependence, but not to identify what causes the difference in the response patterns. Although any stimulus prior to a probe stimulus—i.e., any conditioning stimulus—could in principle provide contextual effects on the neural responses to the probe stimulus (see *Stimulus design*), the mode of context dependence was different from one cell to another and we could not find any particular context–probe combinations that always or never gave an effect. Stimulus space is huge in general and thus we used well-controlled synthetic stimuli to delve into the effects of acoustic properties on the context dependence (see SYNTHETIC SOUNDS).

**Fractional power measure.** A second measure was introduced to examine the contribution of context dependence to response dynamics; i.e., a quantity meaningful from a modeling—instead of just statistical—perspective (see also RESPONSE PREDICTABILITY). In short, we assumed an additive noise model (Eq. 11), in which stimulus-related (predictable) component of the response power—or the second-order statistics at each sampled time point after probe onset over the population—can be decomposed into context-dependent and -independent fractions (Eqs. 17–19). Then the measure was defined as

the context-dependent fractional power normalized by the predictable response power (Eq. 20).

Formally, we assumed the following additive model for the response to a probe stimulus over time  $t$

$$r_{ij}(t) = \mu(t) + \nu_i(t) + \varepsilon_{ij}(t) \quad (11)$$

That is, the observed response  $r_{ij}(t)$  in the  $i$ th context for the  $j$ th trial consists of independent and identically distributed (i.i.d.) Gaussian noise:  $\varepsilon_{ij}(t) \sim \mathcal{N}[0, \sigma_{\text{noise}}^2]$ —with zero mean and the variance of  $\sigma_{\text{noise}}^2$ —and stimulus-related (predictable) parts, which can be further decomposed into context-dependent and -independent fractions:  $\nu_i(t)$  and  $\mu(t)$ , respectively. Each component can be estimated as

$$\hat{\mu}(t) = \langle \langle r_{ij}(t) \rangle \rangle_i \quad (12)$$

$$\hat{\nu}_i(t) = \langle r_{ij}(t) \rangle_j - \langle \langle r_{ij}(t) \rangle \rangle_i \quad (13)$$

$$\hat{\varepsilon}_{ij}(t) = r_{ij}(t) - \langle r_{ij}(t) \rangle_j \quad (14)$$

Then the measure was defined as the context-dependent fractional power:  $\mathcal{P}[\nu_i(t)] = \langle \langle \nu_i^2(t) \rangle \rangle_i$ , normalized by the predictable response power:  $\mathcal{P}[\mu(t) + \nu_i(t)]$  (see Eqs. 20 and 43 in RESPONSE PREDICTABILITY). The “power” is usually computed as the average over time, but here we assumed ergodicity and thus took the average—at each sampled time point after probe onset—over the populations (indicated by the angle brackets without subscripts). In the infinite data limit (i.e., every possible probe examined with all possible conditioning stimuli), the predictable response power should become time-invariant:  $\mathcal{P}[\mu(t) + \nu_i(t)] \rightarrow \sigma^2 = \langle \mathcal{P}[\mu(t) + \nu_i(t)] \rangle_n$ , where the symbol “ $\stackrel{\text{e}}{=}$ ” means “equal in expectation.” In practice, however, the population average of the response power was nonstationary over time—typically, with large fluctuation soon after the transition from conditioning to probe stimuli (Fig. 3B, thin black) because of a finite recording length and stimulus design. Moreover, the estimated power was noisier at longer delay because of the variability in the stimulus duration (see NATURAL SOUNDS). For the sake of normalization, we thus smoothed the predictable response power (the denominator of Eqs. 20 and 43; gray line in Fig. 3B, top) by taking the running average over  $[0, 2t]$  at time  $t$  ( $\geq 0$ ), where the window is selected to be symmetric around the time of interest  $t$ , but always nonnegative. Because a reliable estimate of the predictable power over time could not be obtained at the single-cell level due to data limitation, Figs. 2E and 7C show the unnormalized context-dependent response power (or estimated variance; Eq. 19 without population average):  $\langle \hat{\nu}_i^2(t) \rangle_i$ .

The fractional response powers can be estimated by considering the “average power” and the “power of the average,” under the assumption that the additive components in Eq. 11 are all uncorrelated between each other at any given moment (see also Sahani and Linden 2003). Considering the average over trials (for  $j = 1, \dots, m$ ), we have

$$\langle \langle r_{ij}^2(t) \rangle \rangle_i \stackrel{\text{e}}{=} \mathcal{P}[\mu(t) + \nu_i(t)] + \langle \langle \varepsilon_{ij}^2(t) \rangle \rangle_i \quad (15)$$

$$\langle \langle r_{ij}(t) \rangle \rangle_i \stackrel{\text{e}}{=} \mathcal{P}[\mu(t) + \nu_i(t)] + \langle \langle \varepsilon_{ij}(t) \rangle \rangle_i \quad (16)$$

From the central limit theorem, we have:  $\langle \langle \varepsilon_{ij}(t) \rangle \rangle_i \stackrel{\text{e}}{=} \langle \langle \varepsilon_{ij}^2(t) \rangle \rangle_i / m$ , and thus

$$\hat{\mathcal{P}}[\mu(t) + \nu_i(t)] = \left\langle \left\langle \frac{m \langle r_{ij}(t) \rangle_j^2 - \langle r_{ij}^2(t) \rangle_j}{m - 1} \right\rangle \right\rangle_i \quad (17)$$

where  $\langle \cdot \rangle$  (without subscripts) indicates the average over all the tested probe stimuli in the population data. Similarly, considering the average of the trial average  $\langle r_{ij}(t) \rangle_j$  over contexts (conditioning stimuli;  $i = 1, \dots, n$ ), we have

$$\hat{\mathcal{P}}[\mu(t)] = \left\langle \frac{n \langle \langle r_{ij}(t) \rangle \rangle_i^2 - \langle \langle r_{ij}(t) \rangle \rangle_i^2}{n - 1} \right\rangle \quad (18)$$

Therefore from Eqs. 17 and 18, the context-dependent fractional power can be given in expectation as (note that this *estimated* power can be negative; thick black line in Fig. 3B, top)

$$\hat{\mathcal{P}}[\nu_i(t)] = \hat{\mathcal{P}}[\mu(t) + \nu_i(t)] - \hat{\mathcal{P}}[\mu(t)] \quad (19)$$

and we could use the following quantity  $Q(t)$  as a measure of the contribution of context dependence to response dynamics (black line in Fig. 3B, bottom)

$$Q(t) = \frac{\hat{\mathcal{P}}[\nu_i(t)]}{\hat{\mathcal{P}}[\mu(t) + \nu_i(t)]} \quad (20)$$

where the denominator was smoothed by taking the moving average over  $[0, 2t]$  at time  $t$  before computing Eq. 20 (and Eq. 43 in RESPONSE PREDICTABILITY) to obtain better estimates of  $Q(t)$ . The total context-dependent effects can then be well described as the area under the curve of this fractional power measure over time

$$\lim_{t \rightarrow \infty} \int_0^t Q(s) ds \quad (21)$$

In Fig. 8C, however, the area was computed only for large enough time ( $t = 4$  s) from the population results examined with various stimulus ensembles (Fig. 8B) and normalized by the corresponding area for the effects examined with natural sounds (Fig. 3B). Confidence intervals were computed by resampling methods (200 repeats with randomly selected 1,000 samples).

This fractional power measure in Eq. 20 differs from the significance measure in two ways. First, the fractional power measure is continuous over time, whereas the significance measure involves arbitrary thresholding procedure to determine the significance level and is consequently binary. Second, the fractional power measure involves the normalization (by the stimulus-related response power) and thus depends on the “relative” difference to the overall fluctuation between the response patterns caused by stimulus history and its context, whereas the significance measure (and the numerator of the fractional power measure, i.e., the context-dependent fractional power in Eq. 19) depends on the “absolute” observed differences. Therefore the fractional power measure would be more reliable in these respects (see also the relation to the response predictability; RESPONSE PREDICTABILITY).

CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL. For those recordings for which we could test at least two repeats of context–probe sequences at all three holding potentials, we first estimated net evoked synaptic conductance and computed its excitatory and inhibitory components as described previously (Monier et al. 2008; Wehr and Zador 2003, 2005). As a measure of context dependence, we then computed the variance of the conductances (evoked by probe stimuli) over contexts and took the population average (over 29 probes tested in 14 cells; Fig. 6; see also Eqs. 30 and 31).

*Estimation of evoked synaptic conductance.* To compute total synaptic conductance,  $g_{\text{syn}}(t)$ , a model of a linear, isopotential neuron was used (see also Monier et al. 2008; Wehr and Zador 2003, 2005). The total synaptic current  $I_{\text{syn}}(t)$  is then given by the following generic membrane equation

$$I_{\text{syn}}(t) = g_{\text{syn}}(t)[V(t) - E_{\text{syn}}(t)] \quad (22)$$

where  $E_{\text{syn}}(t)$  is the net synaptic reversal potential and  $V(t)$  is the holding potential corrected off-line for somatic voltage escape and a liquid junction potential ( $V_{\text{jp}}$ , calculated to be  $\sim 12$  mV; Barry 1994). That is

$$V(t) = V_{\text{hold}} - I(t)R_s - V_{\text{jp}} \quad (23)$$

where  $V_{\text{hold}}$  is the command potential,  $R_s$  is the series resistance, and  $I(t)$  is the total recorded current. In this study  $R_s$  was computed from the peak current transients by taking the median across the train of 10 voltage pulses (preceding acoustic stimuli), and no on-line series resistance compensation was used (see *Electrophysiology*). The total current  $I(t)$  is given by the sum of the synaptic current  $I_{\text{syn}}(t)$  and the additional holding current required to voltage-clamp the soma away from its resting potential  $E_{\text{rest}}$

$$I(t) = I_{\text{syn}}(t) + \frac{V(t) - E_{\text{rest}}}{R_{\text{in}}} \quad (24)$$

where  $R_{\text{in}}$  is the input resistance estimated from the steady-state current induced by the voltage pulses. Note that  $E_{\text{rest}}$  can also be estimated by the steady-state current–voltage ( $I$ – $V$ ) relationship, where  $I_{\text{syn}}(t)$  is assumed to be zero. From Eqs. 23 and 24, the sound-evoked synaptic current can be specifically expressed in terms of the change in the recorded currents  $\Delta I(t)$ , due to stimulus presentation

$$I_{\text{syn}}(t) = \frac{R_{\text{in}} + R_s}{R_{\text{in}}} \Delta I(t) = \frac{R_{\text{in}} + R_s}{R_{\text{in}}} [I(t) - I_{\text{baseline}}] \quad (25)$$

where  $I_{\text{baseline}}$  is the baseline, spontaneous current level for a given holding potential, estimated as the 90 percentile value of the recorded currents for each trial.

Using  $V(t)$  and  $I_{\text{syn}}(t)$  obtained from Eqs. 23 and 24, respectively, we can in principle solve Eq. 22 for  $g_{\text{syn}}(t)$  and  $E_{\text{syn}}(t)$  by linear regression. In practice, however, here we used  $I_{\text{syn}}(t)$  from Eq. 25 to minimize potential distortions introduced by any additional nonlinearities in the steady-state  $I$ – $V$  relationship and by nonstationarity in the recordings. Note that the net evoked synaptic conductance can be negative, meaning a reduction of the conductance relative to the baseline synaptic conductance.

*Conductance decomposition into excitatory and inhibitory components.* The excitatory and inhibitory components of the total synaptic conductance—indicated by the subscripts “exc” and “inh,” respectively—can be extracted by assuming the following conductance model

$$g_{\text{syn}}(t) = g_{\text{exc}}(t) + g_{\text{inh}}(t) \quad (26)$$

$$g_{\text{syn}}(t)E_{\text{syn}}(t) = g_{\text{exc}}(t)E_{\text{exc}} + g_{\text{inh}}(t)E_{\text{inh}} \quad (27)$$

Solving Eqs. 26 and 27 for  $g_{\text{exc}}(t)$  and  $g_{\text{inh}}(t)$ , we have

$$g_{\text{exc}}(t) = g_{\text{syn}}(t) \frac{E_{\text{syn}}(t) - E_{\text{inh}}}{E_{\text{exc}} - E_{\text{inh}}} \quad (28)$$

$$g_{\text{inh}}(t) = g_{\text{syn}}(t) \frac{E_{\text{exc}} - E_{\text{syn}}(t)}{E_{\text{exc}} - E_{\text{inh}}} \quad (29)$$

Here we set  $E_{\text{exc}} = 0$  mV and  $E_{\text{inh}} = -85$  mV by our internal solution (see *Electrophysiology* in preceding text).

*Context-dependence measurement.* For those recordings that we could test at least two repeats of context–probe sequences at all three holding potentials, we first estimated evoked synaptic conductance,  $\hat{g}_{\text{syn}}^i(t)$ , using regression in Eq. 22 for each conditioning stimulus  $i = 1, \dots, n$ , and computed its excitatory and inhibitory components,  $\hat{g}_{\text{exc}}^i(t)$  and  $\hat{g}_{\text{inh}}^i(t)$ , from Eqs. 28 and 29, respectively. As a measure of context dependence, we then estimated the variance of true conductances  $g_{\#}^i(t)$  over contexts,  $\sigma_{g_{\#}^i}^2(t)$  for  $t \geq 0$ , where  $\#$  is either “syn,” “exc,” or “inh.” Assuming i.i.d. noise on conductance estimation,  $\varepsilon_{\#}^i(t) = \hat{g}_{\#}^i(t) - g_{\#}^i(t)$ , in expectation we have

$$\sigma_{g_{\#}^i}^2(t) \stackrel{\varepsilon}{=} \sigma_{g_{\#}^i}^2(t) - \sigma_{\varepsilon_{\#}^i}^2, \quad (30)$$

where  $\sigma_{g_{\#}^i}^2(t)$  is the variance of the estimated conductances, estimated here as the population average (over 29 probes tested in 14 cells)

$$\sigma_{g_{\#}^i}^2(t) = \left\langle \frac{1}{n-1} \sum_i [\hat{g}_{\#}^i(t) - \langle \hat{g}_{\#}^i(t) \rangle_i]^2 \right\rangle \quad (31)$$

and  $\sigma_{\varepsilon_{\#}^i}^2$  is the variance of the noise, computed using the conductance estimate on the spontaneous activity in the absence of stimuli (averaged over time and population).

*Estimation errors.* For conductance estimation in Eqs. 28 and 29, we calculated the excitatory and inhibitory reversal potentials by our internal solution. However, errors in these estimates— $\Delta E_{\text{exc}}$  and  $\Delta E_{\text{inh}}$ , respectively—result in the errors in the conductance decomposition by

$$\Delta g_{\text{exc}}(t) \equiv \hat{g}_{\text{exc}}(t) - g_{\text{exc}}(t) = -g_{\text{exc}}(t) \times \frac{E_{\text{exc}}\Delta E_{\text{inh}} - E_{\text{inh}}\Delta E_{\text{exc}} + (\Delta E_{\text{exc}} - \Delta E_{\text{inh}})E_{\text{syn}}(t)}{[E_{\text{syn}}(t) - E_{\text{inh}}](E_{\text{exc}} - E_{\text{inh}} + \Delta E_{\text{exc}} - \Delta E_{\text{inh}})} \quad (32)$$

$$\Delta g_{\text{inh}}(t) \equiv \hat{g}_{\text{inh}}(t) - g_{\text{inh}}(t) = g_{\text{inh}}(t) \times \frac{E_{\text{exc}}\Delta E_{\text{inh}} - E_{\text{inh}}\Delta E_{\text{exc}} + (\Delta E_{\text{exc}} - \Delta E_{\text{inh}})E_{\text{syn}}(t)}{[E_{\text{exc}} - E_{\text{syn}}(t)](E_{\text{exc}} - E_{\text{inh}} + \Delta E_{\text{exc}} - \Delta E_{\text{inh}})} \quad (33)$$

In particular, erroneous estimate of the holding potential  $V(t)$  leads to  $\Delta E_{\text{exc}} = \Delta E_{\text{inh}} (= \Delta E)$  and thus the estimation errors for excitatory and inhibitory conductances are given by

$$\Delta g_{\text{exc}}(t) = -g_{\text{exc}}(t) \frac{\Delta E}{E_{\text{syn}}(t) - E_{\text{inh}}} \quad (34)$$

$$\Delta g_{\text{inh}}(t) = g_{\text{inh}}(t) \frac{\Delta E}{E_{\text{exc}} - E_{\text{syn}}(t)} \quad (35)$$

The estimated strength of context dependence (see Eqs. 30 and 31) will then be in error. However, the estimated timescale will be less affected because the errors in Eqs. 34 and 35 do not substantially affect the temporal pattern of the estimated conductances (see following text).

*Estimation bias.* In this study we assumed a linear, isopotential neuron to estimate synaptic conductance from voltage-clamp measurements. However, such an assumption does not hold in reality and thus our estimates will contain certain systematic errors (in addition to the errors described earlier). As described previously (see supplemental information in Wehr and Zador 2003), these errors result in underestimates of the absolute conductance and biased decomposition into excitatory and inhibitory components, but do not significantly affect estimates of relative timing. Therefore the strength of context dependence in the excitatory and inhibitory conductances can be underestimated, although the timescale will be less affected by the bias inherent to our estimation methods. Importantly, the emergence and detectability of the context dependence per se should be less affected, specifically because responses were compared across different conditioning stimuli within a given cell.

Here we analytically discuss the effects of cable attenuation (“space-clamp” problem) on estimating synaptic conductance and context dependence. In the regime of small synaptic conductances, the estimated net synaptic conductance  $\hat{g}_{\text{syn}}$  and the estimated synaptic reversal potential  $\hat{E}_{\text{syn}}$  underestimate the true  $g_{\text{syn}}$  and  $E_{\text{syn}}$  by

$$\hat{g}_{\text{syn}} = \delta^2 g_{\text{syn}} \quad (36)$$

$$\hat{E}_{\text{syn}} = E_{\text{rest}} + \frac{E_{\text{syn}} - E_{\text{rest}}}{\delta} \quad (37)$$

where  $\delta \leq 1$  is the electrotonic voltage attenuation factor between soma and dendrite and  $E_{\text{rest}}$  is the resting potential (Carnevale and Johnston 1982; Koch et al. 1982; Zador et al. 1995). Together with Eqs. 28 and 29, we then have underestimates of excitatory and inhibitory conductances as

$$\hat{g}_{\text{exc}} = \delta \left[ \delta + (1 - \delta) \frac{E_{\text{syn}} - E_{\text{rest}}}{E_{\text{syn}} - E_{\text{inh}}} \right] g_{\text{exc}} = \delta \left[ 1 - (1 - \delta) \frac{E_{\text{inh}} - E_{\text{rest}}}{E_{\text{inh}} - E_{\text{syn}}} \right] g_{\text{exc}} \quad (38)$$

$$\hat{g}_{\text{inh}} = \delta \left[ \delta + (1 - \delta) \frac{E_{\text{syn}} - E_{\text{rest}}}{E_{\text{syn}} - E_{\text{exc}}} \right] g_{\text{inh}} = \delta \left[ 1 - (1 - \delta) \frac{E_{\text{exc}} - E_{\text{rest}}}{E_{\text{exc}} - E_{\text{syn}}} \right] g_{\text{inh}} \quad (39)$$

Therefore cable attenuation causes a bias in the decomposition as follows

$$\frac{\hat{g}_{\text{exc}}}{g_{\text{exc}}} \geq \frac{\hat{g}_{\text{inh}}}{g_{\text{inh}}} \quad \text{if } E_{\text{exc}} > E_{\text{syn}} > E_{\text{rest}} > E_{\text{inh}} \quad \text{and} \\ \hat{g}_{\text{inh}} \leq 0 \quad \text{if } \delta \leq \frac{E_{\text{syn}} - E_{\text{rest}}}{E_{\text{exc}} - E_{\text{rest}}} \quad \text{while } \hat{g}_{\text{exc}} \geq 0 \quad \text{for all } \delta \leq 1 \quad (40)$$

$$\frac{\hat{g}_{\text{inh}}}{g_{\text{inh}}} \geq \frac{\hat{g}_{\text{exc}}}{g_{\text{exc}}} \quad \text{if } E_{\text{exc}} > E_{\text{rest}} > E_{\text{syn}} > E_{\text{inh}} \quad \text{and} \\ \hat{g}_{\text{exc}} \leq 0 \quad \text{if } \delta \leq \frac{E_{\text{rest}} - E_{\text{syn}}}{E_{\text{rest}} - E_{\text{inh}}} \quad \text{while } \hat{g}_{\text{inh}} \geq 0 \quad \text{for all } \delta \leq 1 \quad (41)$$

In our recording data, the evoked synaptic reversal potential was mostly above rest ( $\hat{E}_{\text{syn}} > E_{\text{rest}}$ ). From Eq. 37, the true synaptic reversal potential was then likely above rest ( $E_{\text{syn}} > E_{\text{rest}}$ ) and thus the inhibitory conductance was relatively more underestimated (so as the ratio of inhibitory to excitatory conductance), as shown in Eq. 40.

Despite the bias in the relative magnitudes of the excitatory and inhibitory components, the temporal features of these components are much less affected by cable attenuation if the change in  $g_{\text{exc}}$  (or  $g_{\text{inh}}$ ) over time is relatively larger than that in the attenuation ratio  $\hat{g}_{\text{exc}}/g_{\text{exc}}$  (or  $\hat{g}_{\text{inh}}/g_{\text{inh}}$ , due mostly to the change in  $E_{\text{syn}}$ ). Moreover, although context-dependent changes in  $E_{\text{syn}}$  can affect the context dependence in both  $\hat{g}_{\text{exc}}$  and  $\hat{g}_{\text{inh}}$ , such effects will be small because  $\partial \hat{g}_{\text{exc}}/\partial E_{\text{syn}} \sim \mathcal{O}(g_{\text{exc}}/E_{\text{syn}}^2)$  and  $\partial \hat{g}_{\text{inh}}/\partial E_{\text{syn}} \sim \mathcal{O}(g_{\text{inh}}/E_{\text{syn}}^2)$  from Eqs. 38 and 39, respectively, and the other variables involved in the estimation bias— $\delta$ ,  $E_{\text{exc}}$ ,  $E_{\text{inh}}$ , and  $E_{\text{rest}}$ —will be less likely to show substantial context dependence. Therefore our ability to detect context dependence in excitatory and inhibitory conductance should be less deteriorated, and the estimated timescale of context dependence will be more likely to be faithful.

**CONTEXT DEPENDENCE AT SUPRATHRESHOLD LEVEL.** Context dependence at the spike level (in thalamus) was analyzed for those recordings that we could test  $\geq 10$  repeats of any given combinations of a probe stimulus and at least two conditioning stimuli (Fig. 9). Using the bin size of  $\Delta t = 100$  ms, we first generated poststimulus time histograms  $q_i(t_k)$  for all the conditioning stimuli  $i = 1, \dots, n$  we examined, where  $\Delta t \cdot k < t_k \leq \Delta t \cdot (k + 1)$  ms for  $k = 0, 1, \dots, 60$

(see, e.g., Fig. 9B). As a measure of context dependence, we then computed the SD of  $q_i(t_k)$  over contexts (Fig. 9C) and took the average across the population (over all 93 probes tested in 14 thalamic neurons; Fig. 9D)

$$\left\langle \sqrt{\frac{1}{n-1} \sum_i (q_i(t_k) - \langle q_i(t_k) \rangle)^2} \right\rangle \quad (42)$$

**RESPONSE PREDICTABILITY.** To analyze how response predictability in A1 depends on stimulus history and its context over time, we computed the time course of the ratio between context-independent fractional power  $\mathcal{P}[\mu(t)]$  and the stimulus-related response power  $\mathcal{P}[\mu(t) + \nu_i(t)]$  (Fig. 4). From a modeling perspective,  $\mathcal{P}[\mu(t)]$  represents the power we could capture in the response estimation exploiting the stimulus history for a limited duration of  $t$  (i.e., using a temporal window from the probe onset), whereas  $\mathcal{P}[\mu(t) + \nu_i(t)]$  gives the upper-bound that no model could outweigh under the additive noise assumption because it uses the entire stimulus history. The context-dependent response power  $\mathcal{P}[\nu_i(t)]$  indicates the fraction that is *not* accessible when only a finite stimulus history (i.e., information only on the “probes”) is available, and that the trial-to-trial noise power  $\sigma_{\text{noise}}^2 \stackrel{\text{§}}{=} \langle \hat{\mathcal{P}}[\varepsilon_{ij}(t)] \rangle$  is the fraction that is *never* accessible under the additive noise assumption. Therefore the following ratio (at time  $t$  after probe onset) indicates the context dependence of the response predictability, which constitutes an upper-bound estimate of the response prediction performance for a given window length  $t$

$$\frac{\hat{\mathcal{P}}[\mu(t)]}{\hat{\mathcal{P}}[\mu(t) + \nu_i(t)]} \quad (43)$$

where the denominator and numerator were computed from Eqs. 17 and 18, respectively.

Two points should be mentioned here. First, Eq. 43 also indicates (the population average of) the fraction of the stimulus-related power attributable only to the probe stimulus at time  $t$  after probe onset. Second, it is the context-independent fraction  $\mathcal{P}[\mu(t)]$  that characterizes the model performance (Fig. 4), whereas it is the context-dependent fraction  $\mathcal{P}[\nu_i(t)]$  that we used to characterize the neuronal behaviors (Figs. 2, 3, 7, and 8).

**EXPONENTIAL CURVE FIT.** To measure the relevant timescales of the context dependence, we fit (a sum of) exponential processes to the population data (Figs. 3, 4, 8, 9D, and Supplemental Fig. S1)

$$\alpha + \sum_k \alpha_k \exp\left[-\frac{t}{\tau_k}\right] \quad (44)$$

where  $\alpha_k$  and  $\tau_k$  indicate the decay size and constant, respectively. We used the `lsqcurvefit` function in MATLAB Optimization Toolbox for the curve fitting and used the following criterion for choosing the number of exponential processes:  $|\alpha_k| > \sum_k |\alpha_k|/10$  for all  $k$ , that is, the contribution of an exponential process must be at least one tenth of the total.

**TIMESCALE OF INTRINSIC MEMBRANE PROPERTIES.** The noise correlation was computed as the autocorrelation of additive noise  $\varepsilon_{ij}(t)$  (defined as in Eq. 11 and estimated as in Eq. 14)

$$\vartheta_{ij}(t) = \frac{\varepsilon_{ij}(-t) * \varepsilon_{ij}(t)}{\sigma_{\text{noise}}^2} \quad (45)$$

where  $*$  indicates convolution. The correlation function  $\vartheta_{ij}(t)$  represents a similarity in the trial-to-trial noise components  $\varepsilon_{ij}(t)$  over time and thus characterizes the timescale of the intrinsic (stimulus-independent) dynamics of the membrane potential. However, we found that  $\vartheta_{ij}(t)$  had a sharp peak with a rapid decay (within  $\sim 100$  ms; data not shown), much faster than the slow timescale identified in neural responses in area A1 (Fig. 3). The autocorrelation of the spontaneous

activity was computed in a similar manner using Eq. 45, which also decayed within roughly 100 ms (data not shown).

**NEURAL ENCODING MODELS.** We used linear–nonlinear cascade models (Klein et al. 2000; Machens et al. 2004) and compared their performance to the upper-bound estimate (Eq. 43) for further analyzing the context dependence of the response predictability (Fig. 4). The estimated response by a linear spectrotemporal receptive field (STRF) model,  $\hat{r}(t)$ , is given by

$$\hat{r}(t) = \iint \text{STRF}(\tau, \omega) \cdot S(t - \tau, \omega) d\tau d\omega \quad (46)$$

where  $S(t, \omega)$  is the spectrogram (short-time Fourier transform) of the sound pressure waveform  $s(t)$

$$S(t, \omega) = 20 \log_{10} \left| \frac{1}{\sqrt{2\pi}} \int e^{-j\omega\tau} s(\tau) h(\tau - t) d\tau \right| \quad (47)$$

where  $h(t)$  is a window function (Cohen 1995). Because the recording data we collected for examining context dependence were not tested with enough varieties of stimuli (see *Stimulus design*), which could cause a bias in the STRF estimation (Paninski 2003; Simoncelli et al. 2004), here we (re)analyzed the recording data (20 cells) from the previous work (Machens et al. 2004).

**Parameter estimation.** We used the ridge regression technique to obtain the best estimate of STRF, as detailed in Hastie et al. (2001) (in the context of neuroscience, see, e.g., Machens et al. 2004; Wu et al. 2006). In short, we discretized time  $t$  and frequency  $\omega$  and reordered the indices to simplify Eq. 46 into the following form

$$\hat{\mathbf{r}} = \mathbf{S}\boldsymbol{\beta} \quad (48)$$

where  $\hat{\mathbf{r}}$  and  $\boldsymbol{\beta}$  are column vectors of the estimated response and the STRF, respectively, and the  $i$ th row of the matrix  $\mathbf{S}$  consists of the  $i$ th stimulus vector. (Here we use **boldface** to indicate vectors and matrices in lower- and uppercase letters, respectively.) Ridge regression is one of the shrinkage methods to penalize strong deviations of the parameters from zero, and the error (objective) function to be minimized is given as

$$E_{\text{ridge}}(\boldsymbol{\beta}, \lambda) = \|\mathbf{r} - \mathbf{S}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|^2 \quad (49)$$

where  $\|\cdot\|$  indicates  $L_2$ -norm and the parameter  $\lambda \geq 0$  determines the strength of the ridge (power) constraint. The solution that minimizes Eq. 49 is then given as

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{S}^T\mathbf{S} + \lambda\mathbf{I})^{-1}\mathbf{S}^T\mathbf{r} = \mathbf{V}(\boldsymbol{\Sigma}^2 + \lambda\mathbf{I})^{-1}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{r} \quad (50)$$

where  $\mathbf{I}$  is the identity matrix,  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal matrices whose columns span the column space of  $\mathbf{S}$  and  $\mathbf{S}^T$ , respectively, and  $\boldsymbol{\Sigma}$  is a diagonal matrix of the singular values given by the singular value decomposition:  $\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ . (Here we used superscript “ $T$ ” to indicate a matrix transpose and used the svd function in MATLAB for the computation.)

In this study we performed 10-fold cross-validation, i.e., split the data set into training (90%) and validation (10%) data sets, used the training data set to estimate STRF as in Eq. 50 with various values of  $\lambda$ , and chose such  $\lambda$  and corresponding STRF  $\hat{\boldsymbol{\beta}}_{\text{ridge}}$  that gave the best model performance on the validation data set (see Eq. 53 in MODEL PERFORMANCE). The resulting model performance on the training and validation data set can then be considered as the upper and lower estimates, respectively (Ahrens et al. 2008; Machens et al. 2004; Sahani and Linden 2003).

To vary the window length while fixing the model complexity—i.e., the number of free parameters in a model—for a fair comparison of the model performance, we varied the time-bin sizes in a pseudo-logarithmic scale:  $\Delta_k t = 2^k$  ms for  $k = 2, \dots, 10$  from near to distant

past. In this study we set the number of bins for  $\Delta_k t$  (45 in total) as [45, 0,  $\dots$ , 0], [9, 36,  $\dots$ , 0],  $\dots$ , [9, 8,  $\dots$ , 1], resulting in models with window lengths of 180, 324, 548, 884, 1,364, 2,004, 2,772, 3,540, and 4,052 ms, respectively. Frequency discretization was  $\Delta x = 3$  bins/octave (ranging from 0.40 to 22 kHz; 17 frequency bins), leading to 765 parameters in total for the linear part of the model (i.e., STRF in Eq. 46). Static nonlinearities were then identified using a scatterplot between actual responses and the estimates by the STRF (see following text).

**Static nonlinearities.** Static nonlinearities can be given as a nonlinear transformation  $\mathcal{G}_{\text{sn}}$  that acts on the output of the linear model (Eq. 46) to form a new (better) estimate (Machens et al. 2004; Simoncelli et al. 2004)

$$\mathcal{G}_{\text{sn}}: \hat{r}(t) \mapsto \hat{r}_{\text{sn}}(t) \quad (51)$$

For fitting the static nonlinearities, we plotted the actual response  $\mathbf{r}$  against the estimated response  $\hat{\mathbf{r}}_{\text{ridge}} = \mathbf{S}\hat{\boldsymbol{\beta}}_{\text{ridge}}$  from Eqs. 48–50 and used the robust locally weighted scatterplot smoothing (Cleveland 1979; Cleveland and Devlin 1988; Hastie et al. 2001) with 5% data span and five iterations. Here we identified such a continuous transformation  $\mathcal{G}_{\text{sn}}$ , using the training set, and then applied  $\mathcal{G}_{\text{sn}}$  to the validation data set, resulting in upper and lower estimates of the model performance, respectively.

**MODEL PERFORMANCE.** The model performance for Eqs. 46 and 51 was quantified as the ratio between the estimated response power captured by a model  $\sigma_{\text{model}}^2$  and the stimulus-related (predictable) response power  $\sigma^2$  (Ahrens et al. 2008; Machens et al. 2004; Sahani and Linden 2003). Note the similarity to the analysis of response predictability (RESPONSE PREDICTABILITY).

Assuming additive i.i.d. Gaussian noise  $\varepsilon_j(t) \sim \mathcal{N}[0, \sigma_{\text{noise}}^2]$  over trials (for  $j = 1, \dots, m$ ) and time  $t$ , we can express the observed response for the  $j$ th trial as  $r_j(t) = \rho(t) + \varepsilon_j(t)$ , with the stimulus-related components  $\rho(t)$  [equivalent to  $\mu(t) + \nu_j(t)$  in Eq. 11] and the total power in the observed response as  $\sigma_{\text{total}}^2 = \sigma^2 + \sigma_{\text{noise}}^2$ , with the stimulus-related power  $\sigma^2 \equiv \langle \rho^2(t) \rangle_t$  in the limit of large  $t$ . (As before, we use  $\langle \cdot \rangle$  to indicate the average over time with subscript  $t$  and over trials with subscript  $j$  and we have  $\langle \rho(t) \rangle_t = 0$  because  $r_j(t)$  is preprocessed to have zero mean.) From the central limit theorem, the power of the average response over trials can be written as:  $\langle \langle r_j(t) \rangle_j^2 \rangle_t \stackrel{\varepsilon}{=} \sigma^2 + \sigma_{\text{noise}}^2/m$ . Therefore the predictable response power  $\sigma^2$  can be estimated as

$$\hat{\sigma}^2 = \left\langle \frac{m \langle r_j(t) \rangle_j^2 - \langle r_j^2(t) \rangle_j}{m - 1} \right\rangle_t \quad (52)$$

where we use  $\hat{\sigma}_{\text{total}}^2 = \langle \langle r_j^2(t) \rangle_j \rangle_t$ . Note the similarity to Eqs. 17 and 18.

The model performance  $\sigma_{\text{model}}^2/\sigma^2$  is then given as

$$\frac{\hat{\sigma}_{\text{total}}^2 - \hat{\sigma}_{\text{error}}^2}{\hat{\sigma}^2} \quad (53)$$

where  $\hat{\sigma}_{\text{error}}^2 = \langle \langle [r_j(t) - \hat{r}_j(t)]^2 \rangle_j \rangle_t$  is the model error power. In Eq. 43, the average was taken over the population (but not over the time), but the quantities in Eqs. 43 and 53 are equivalent under the assumption of ergodicity.

## RESULTS

We developed a novel experimental paradigm for estimating the time course and magnitude of context-dependent effects on neural responses in rat primary auditory cortex (area A1). We probed neurons with a variety of spectrotemporally rich stimuli (e.g., animal vocalizations) in sequence (see *Stimulus design* in METHODS and Fig. 1). The use of such complex stimuli allowed us to probe a larger fraction of

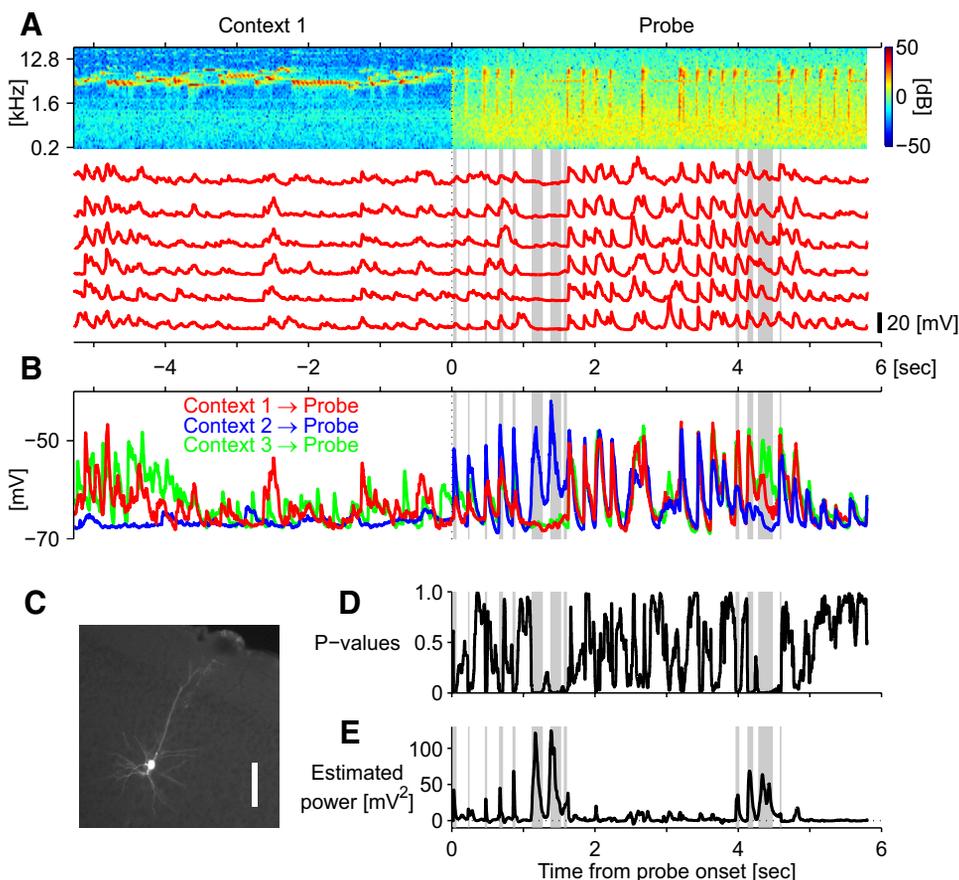
stimulus space than conventional protocols using tones and other simple stimuli (see also Bar-Yosef et al. 2002; Garcia-Lazaro et al. 2006; Machens et al. 2004; Theunissen et al. 2001).

Our analysis consisted of the following five parts. First, we assessed the overall context dependence of neurons in A1 using natural sound ensembles. Second, we quantified the context dependence from the viewpoint of model construction, i.e., measured the response predictability given all the past stimulus information within an arbitrary window length. Third, we examined the context dependence at the synaptic input level by decomposing the acoustic responses into their underlying excitatory and inhibitory components. Fourth, we used synthetic sounds to characterize how context dependence depended on stimulus properties such as stimulus intensity and modulation rates. Finally, we examined thalamic contributions to the context dependence in auditory cortical neurons.

### Context dependence

Firing rates in A1 were typically low under our experimental conditions (spontaneous,  $0.47 \pm 0.61$  Hz; evoked,  $0.57 \pm 0.77$  Hz; mean  $\pm$  SD in 194 cells; see also Hromádka et al. 2008; Wehr and Zador 2005). We therefore examined subthreshold responses rather than firing rates. Because subthreshold responses consist of a continuous variable in time (membrane potential) rather than a sparse binary time series (a train of action potentials), we could obtain good estimates of activity even in the complete absence of spiking outputs. From a modeling perspective subthreshold responses may offer an additional advantage in that they have been subjected to one fewer nonlinearity—that imposed by the spike-generation mechanism—and so may be more linearly related to the stimulus.

Figure 2 shows a typical example of subthreshold responses to a 6-s natural sound stimulus in three different natural sound



**FIG. 2.** Context dependence can persist for several seconds. *A*: typical subthreshold responses of a rat A1 neuron to part of a natural sound sequence (spectrogram; *time 0* indicates the transition from conditioning to probe stimuli) over 6 repeats (red lines). Spikes were clipped by a median filter (window length, 10 ms; see also Supplemental Fig. S2). The neuron showed high trial-to-trial reliability (correlation coefficients across trials to this particular probe stimulus,  $0.74 \pm 0.08$ ; and across trials across all natural sound stimuli examined in this cell,  $0.61 \pm 0.07$ ; mean  $\pm$  SD). *B*: the mean responses to the probe stimulus in 3 different contexts: red line for the one shown in *A* and blue and green lines for the one in response to the same probe stimulus but preceded by “silence” and another conditioning stimulus, respectively. Significant dependence on the stimulus history was observed for  $>4$  s (gray bands; see *D* for details), whereas the responses between 2 and 4 s after the onset of the probe stimulus were not affected by the differences in conditioning stimuli. *C*: the recorded cell in this example was histologically identified as a layer II–III pyramidal neuron (scale bar: 100  $\mu$ m). *D*: we performed a pointwise statistical (Kruskal–Wallis) test for equal medians between the responses to the probe stimulus in all different contexts (black line; *P* values) and the gray bands show the time points where the context dependence was statistically significant under the criterion:  $P < 0.01$  for  $\geq 5$  ms. *E*: the response power estimate that depends on the context [black line;  $\langle \hat{v}_i^2(t) \rangle$ ; from Eqs. 17–19 without population average] well represents the magnitude of the context dependence. (Note that the *estimated* power can be negative; see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL in METHODS for details.) The population average of this quantity with and without normalization by the average predictable power is shown in Fig. 3*B* (thick black curves in the *bottom* and *top* panels, respectively).

contexts, i.e., preceded by three different 6-s conditioning stimuli (see also Supplemental Fig. S2). Consistent with previous work (Machens et al. 2004), this neuron showed high trial-to-trial reliability (Fig. 2A) within each set of trials for which the conditioning stimuli were held fixed: the correlation coefficient of the response traces across trials in a given context was  $0.61 \pm 0.07$  (mean  $\pm$  SD) for the seven natural sound fragments tested in this cell. The reliability varied within a given neuron as a function of the stimuli tested and across neurons; the mean correlation coefficient was  $0.31 \pm 0.09$  (mean  $\pm$  SD) over the population.

Changing the conditioning stimulus—i.e., the stimulus context—caused a dramatic change in the response to the probe stimulus (Fig. 2B). In this example, the effects of the context on the response lasted  $>4$  s. Interestingly, context-induced differences could sometimes be intermittent; the three average response traces showed no difference in the interval 2 to 4 s after the onset of the probe, but diverged again after about 4 s.

We used two measures to quantify the differences in the probe stimulus induced by temporal context (see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL in METHODS for details). The first examined whether the differences in the observed traces were statistically significant (Kruskal–Wallis test,  $P < 0.01$  for  $\geq 5$  ms; Fig. 2D), whereas the second method assessed the component of the response power (variance at a

given time) dependent on stimulus history (Eq. 19; Fig. 2E shows the power without the population average). These two measures generally agreed quite well, as can be confirmed by noting that when at least one trace was significantly different from the others (vertical gray strips), the power was typically high.

We found that the total response power tended to be high on average soon after the transition from conditioning to probe stimuli (Fig. 3B, top panel). This increase in response power at the transition could lead to an overestimate of the contribution of context. We therefore also used a normalized version of the second (power based) measure in which we divided the context-dependent response power by the stimulus-related response power (Eq. 20; Fig. 3B, bottom). This compensates for the effects of the nonstationarity of the response at the transition, and thus provides a more conservative measure of the context dependence.

Although context-dependent effects often manifested intermittently in a given cell (as in Fig. 2), across the population these effects showed an orderly monotonic decay (Fig. 3). Of 305 natural sound probe stimuli tested with different—typically around five to eight—natural sound contexts in 39 cells, significant effects were observed in 204 probes (66.9%; Fig. 3A), and about a quarter (23.7%) of the events occurred longer than one second after the onset of a probe stimulus (Fig. 3A).

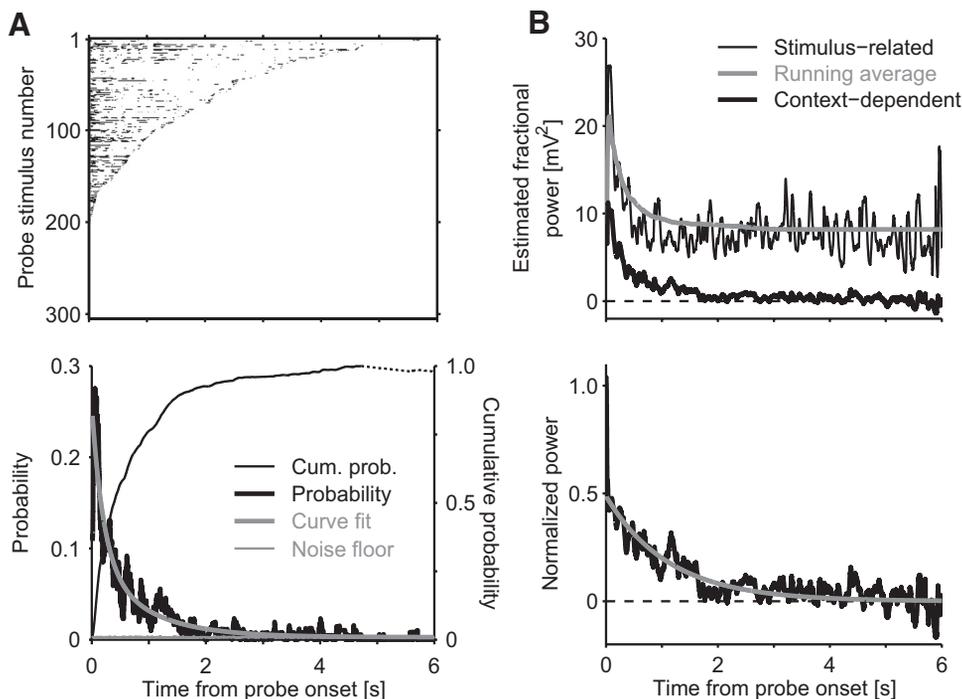


FIG. 3. Long-lasting context dependence in auditory cortex. *A*: significance measure. *Top*: each raster shows periods during which the significance measure exceeds threshold ( $P < 0.01$  for  $\geq 5$  ms; see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL in METHODS) for a particular probe–stimulus combination in a given neuron (an example is shown as gray bands in Fig. 2). The rasters are sorted according to the longest-lasting effect, so successive rasters may correspond to different neurons. Significant context dependence was observed in about two thirds of probe stimuli (204 of 305 probes in 39 neurons). *Bottom*: the thick black curve shows the proportion—or the probability—of observing the significant context dependence and the thin gray curve shows the noise floor computed by resampling methods. The probability curve is well fit by the sum of 2 exponentials (thick gray;  $\alpha_1 = 0.17$ ,  $\tau_1 = 0.20$  s,  $\alpha_2 = 0.09$ , and  $\tau_2 = 0.90$  s for Eq. 44 with the mean noise floor over time  $\alpha = 2.8 \times 10^{-3}$ ). Around a quarter of the context-dependent events occurred at  $\geq 1$  s from the onset of probe stimuli (thin black; cumulative probability corrected for the noise floor and normalized at its peak  $t = 4.72$  s; broken part indicates the events at the noise level). *B*: fractional power measure. *Top*: from the same population data, we computed the stimulus-related response power (thin black;  $\hat{\mathcal{P}}[\mu(t) + \nu_i(t)]$  in Eq. 17; gray, its moving average over the data in  $[0, 2t]$  at time  $t$ ) and the fraction that depends on stimulus history and its context (thick black;  $\hat{\mathcal{P}}[\nu_i(t)]$  in Eq. 19). The context-dependent fraction corresponds well to the significance measure as shown in *A* (bottom). See also Fig. 1A. *Bottom*: the ratio of the context-dependent power to the stimulus-related power represents well the contribution of stimulus history to the response dynamics (black; Eq. 20). The decay size and constant are:  $\alpha_1 = 0.49$ ,  $\tau_1 = 1.04$  s with  $\alpha = 0$  (gray).

This fraction represents a lower bound on the maximum duration of the possible effect in a given cell, since the number of conditioning–probe combinations tested per neuron was quite small, and was not tailored to the properties of the cell. For both measures there was a long decay constant of about one second ( $\tau = 0.90$  and  $1.04$  s, respectively; see Fig. 3 for details). This timescale was much longer than that imposed by the intrinsic membrane properties or the time course of the stimulus-evoked synaptic events ( $\sim 100$  ms; see Eq. 45 in TIMESCALE OF INTRINSIC MEMBRANE PROPERTIES), suggesting that it arose from cortical network rather than single neuron mechanisms (see also *Context dependence of synaptic conductances and Subcortical contribution to context dependence*).

### Relation to response predictability

We have shown that temporal context can influence neuronal responses in area A1 for as long as several seconds. To what extent do these context-dependent effects limit the success of predictive models describing the input–output behavior of A1 neurons? To address this question, we compared the best possible model performance with and without knowledge of stimulus history (see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL and RESPONSE PREDICTABILITY in METHODS for details).

To estimate the best model performance achievable, we assumed that the experimentally observed responses to a given probe stimulus consisted of the sum of a deterministic stimulus-dependent component and a stochastic stimulus-independent (noise) component responsible for trial-to-trial variability. The magnitude of the deterministic component was estimated using methods similar to those introduced in Sahani and Linden (2003) (Eqs. 15–17; see also Eq. 52 in MODEL PERFORMANCE). We then further assumed that the deterministic component could be decomposed into context-dependent and -independent components (see Eq. 11). Under these assumptions, the optimal estimate of the response to a probe stimulus given a particular context is obtained by averaging responses over all presentations of the probe preceded by that context; this is the very best response model achievable under the additive noise assumption. The optimal context-independent estimate is obtained by averaging responses over all presentations of the probe, regardless of the preceding context; this is the best model achievable in the absence of knowledge of the context (i.e., using a temporal window from the probe onset; see also Eqs. 12–14). The context-dependent estimate will inevitably be superior to—or equal to, in the case where context provides no information—the context-independent estimate because it incorporates the effect of the stimulus history.

By comparing the performance of the above-cited two models, we then estimated an upper bound on the best possible prediction achievable from a fixed window (Eq. 43). The estimated upper bound (thick black curve in Fig. 4) shows that no model can capture more than a half ( $1 - |\alpha_1| = 0.51$ ) of the response power given a window length of  $< 100$  ms. To achieve prediction accuracy beyond that, however, stimulus history over seconds must be considered ( $\tau = 1.04$  s, thick gray curve).

This long timescale may explain in part why classical linear encoding (spectrotemporal receptive field [STRF]) models with a limited window length—typically, a few hundred mil-

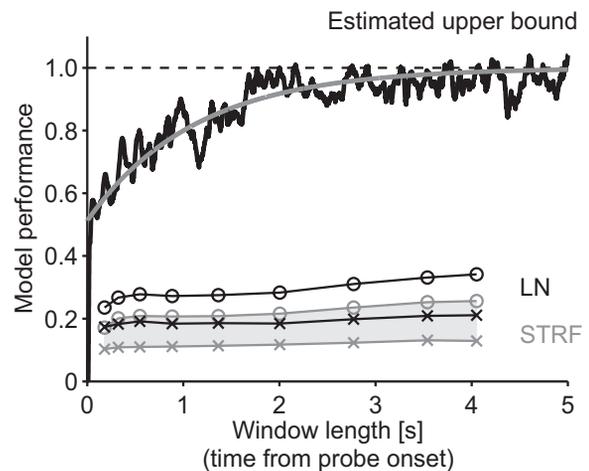


FIG. 4. Context dependence of response predictability. Using responses to natural sound probe stimuli in different natural sound contexts (305 probe stimuli in 39 cells), we computed the ratio between the context-independent fraction of the response power and the stimulus-related response power in A1 neurons (thick black; Eq. 43 in RESPONSE PREDICTABILITY). The stimulus-related fraction is given by the mean response over trials in each context—or the best response estimate under additive noise assumption—whereas the context-independent fraction is given by the mean over all contexts or the best estimate of the responses to a probe stimulus without any knowledge on the conditioning stimulus. Therefore the ratio (at time  $t$  after probe onset) represents the upper bound of the response prediction performance for a given window length  $t$ , which asymptotically approached the true upper limit (black broken line; unit model performance) by extending the window length—or available stimulus history—on the timescale of seconds (thick gray;  $\alpha_1 = -0.49$  and  $\tau_1 = 1.04$  s for exponential curve fit as in Eq. 44 with  $\alpha = 1$ ). In contrast, the performance (Eq. 53; 20 cells from Machens et al. 2004) of linear encoding models (spectrotemporal receptive field [STRF], thin gray; Eq. 46) was low for any window length up to about 4 s, even with static nonlinearities (LN, thin black; Eq. 51). Crosses and open circles, respectively, show the average model performance on the validation and training data sets, corresponding to the lower and upper estimates of the performance. Here we varied the bin sizes in a pseudologarithmic manner for changing the window length of the STRF models while fixing the model complexity (for details, see NEURAL ENCODING MODELS in METHODS).

liseconds—have not provided good predictions for some stimulus ensembles. The performance of STRF-based models was in general unsatisfactory ( $\sim 20\%$ ), consistent with previous work (Ahrens et al. 2008; Machens et al. 2004; Sahani and Linden 2003). The performance did not improve significantly, however, when we extended the window length (up to  $\sim 4$  s; light gray bands in Fig. 4 for mean lower- and upper-bound estimates), even when we added static nonlinearities (thin black lines). This failure could result from inappropriate choices of the model class and/or the initial transformation of sound stimuli from the time domain into the time–frequency domain (Eq. 47 in NEURAL ENCODING MODELS; see also Gill et al. 2006). Instead, it could be simply because we used rather coarse time and frequency resolutions and thus relevant information for the neurons might have been lost. However, we could not identify distinct structures or “features” in the STRFs longer than several hundred milliseconds, suggesting a role of A1 neurons in more than detecting instantaneous stimulus features (Ahrens et al. 2008; Nelken et al. 2003). It is a future challenge to address how neurons in A1 exploit stimulus history and its context on such a long timescale and how we could build a plausible predictive model (see also DISCUSSION, *Context dependence and model construction*).

### Context dependence of synaptic conductances

What are the mechanisms responsible for the long-lasting context dependence in responses of auditory cortical neurons? To address this question, we directly measured sound-evoked synaptic currents by voltage clamping neurons to three different holding potentials and decomposed the responses into their underlying excitatory and inhibitory components (Monier et al. 2008; Wehr and Zador 2003, 2005; see also CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL). Figure 5B shows an example of these synaptic currents in response to a sequence of context-probe stimuli shown in Fig. 5A (see also Supplemental Fig. S3). Evoked synaptic currents were inward at hyperpolarized holding potentials and outward at depolarized holding potentials, consistent with a mixture of excitatory and inhibitory conductances (Fig. 5C; Monier et al. 2008; Wehr and Zador 2003, 2005).

In this cell we also measured the responses to the same probe stimulus following five other conditioning stimuli (Supplemental Figs. S4–S8) and extracted excitatory and inhibitory conductances accordingly (Fig. 5, D and F, respectively). In all cases we examined, both excitatory and inhibitory conductances elicited in this cell decayed rapidly within about 100 ms, yet the context dependence measured as the variance of the conductances lasted long, especially for the inhibitory conductance in this example (Fig. 5E for excitation and Fig. 5G for inhibition). This suggests that a change of synaptic input, rather than a persistent inhibitory or excitatory current, is responsible for the context dependence in responses of auditory cortical neurons.

Although context-dependent effects often appeared differently between excitatory and inhibitory components of the responses in a given cell (as in Fig. 5), across the population these effects showed monotonic decays for both excitation and inhibition (29 probes tested in 14 cells; Fig. 6, A and B, respectively). For both excitatory and inhibitory conductances, there was a long decay constant on the order of seconds ( $\tau = 1.22$  and  $2.50$  s, respectively) and a short decay constant on the order of hundreds of milliseconds ( $\tau = 0.14$  and  $0.13$  s, respectively). These time constants are comparable to those we measured by current-clamping neurons (see *Context dependence* and Fig. 3), even though here we did not compensate for the nonstationarity effects of the responses, and also similar to those reported for activity-dependent, short-term synaptic plasticity (Abbott et al. 1997; Tsodyks and Markram 1997; Wehr and Zador 2005; Zucker and Regehr 2002). We therefore conclude that the context dependence described is either inherited from thalamic inputs—which we reject later in *Subcortical contribution to context dependence* in the following text—or generated by synaptic depression and/or facilitation at thalamocortical or intracortical synapses.

### Relation to stimulus properties

Thus far we used natural sounds because of their rich spectrotemporal structure and because the ultimate test of a model is whether it is able to account for responses to arbitrary stimuli. However, a disadvantage of using natural sounds as stimuli is that we could not readily determine which stimulus properties were responsible for the long-lasting context effects

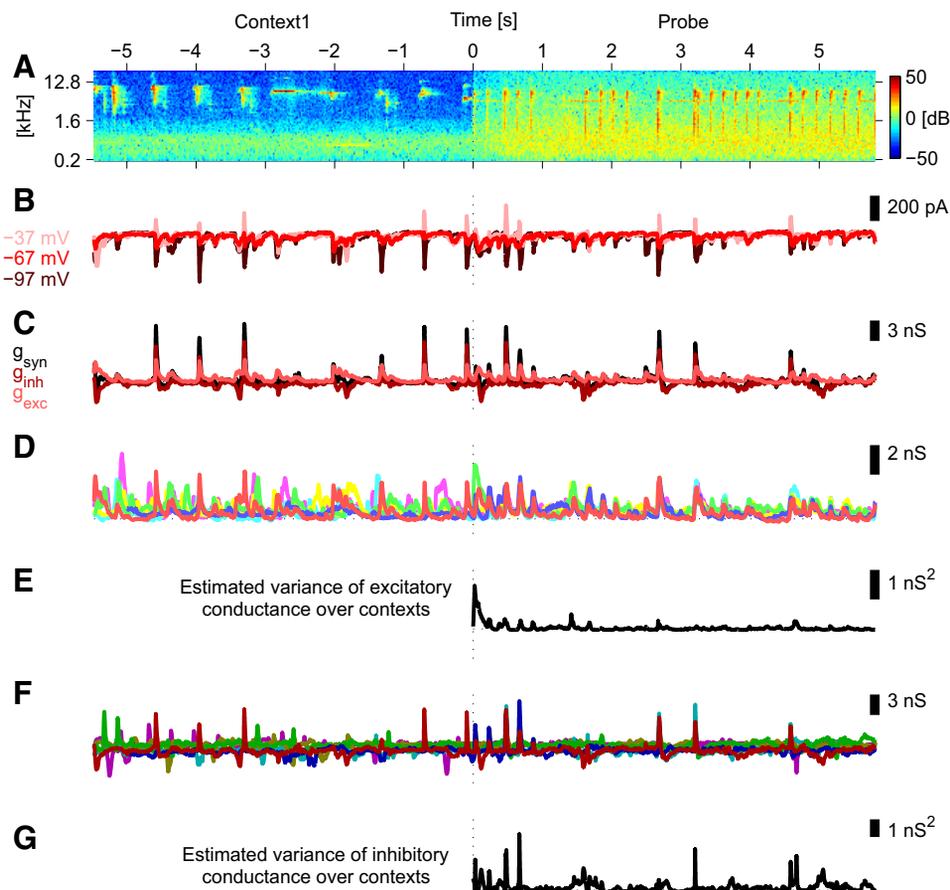


FIG. 5. Excitatory and inhibitory conductances showed different context dependence patterns. *A*: spectrogram of part of a natural sound sequence presented to a rat A1 neuron. Time zero indicates the transition from conditioning to probe stimuli. *B*: example of the mean sound-evoked synaptic currents at 3 different holding potentials. The raw traces are shown in Supplemental Fig. S3B. Spikes were blocked pharmacologically. *C*: estimated total synaptic conductance ( $g_{\text{syn}}$  in Eqs. 26 and 27 in CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL in METHODS) and its excitatory and inhibitory components ( $g_{\text{exc}}$  and  $g_{\text{inh}}$  from Eqs. 28 and 29, respectively) evoked by natural sounds shown in *A*. See also Supplemental Fig. S3C. Note that both evoked excitatory and inhibitory conductances typically decayed within about 100 ms. *D* and *F*: estimated excitatory and inhibitory conductances in response to the probe stimulus in 6 different contexts (*D* and *F*, respectively). Estimates for different contexts are represented by different colors (red lines for those shown in *C*; see also *C* in Supplemental Figs. S3–S8) and estimated excitatory and inhibitory conductances are shown in light and dark colors, respectively. *E* and *G*: estimated variance of excitatory and inhibitory conductances over the 6 contexts for the probe period (*E* and *G*, respectively; Eq. 30 without population average). In this example, context dependence in inhibitory conductance was longer than that in excitatory conductance.

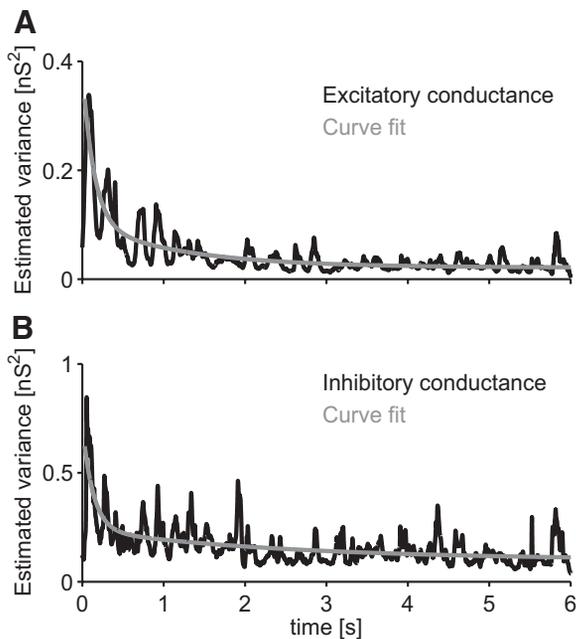


FIG. 6. Context dependence of excitatory and inhibitory conductances. *A*: the population average of the variance of excitatory conductance over contexts (black; 29 probes tested in 14 cells; Eq. 30 in CONTEXT DEPENDENCE AT SYNAPTIC INPUT LEVEL in METHODS). The data were well fit by the sum of 2 exponentials (gray;  $\alpha_1 = 0.28$ ,  $\tau_1 = 0.14$  s,  $\alpha_2 = 0.08$ , and  $\tau_2 = 1.22$  s with  $\alpha = 0.02$  in Eq. 44). *B*: the population average of the variance of inhibitory conductance over contexts (black; 29 probes tested in 14 cells). The data were well fit by the sum of 2 exponentials (gray;  $\alpha_1 = 0.48$ ,  $\tau_1 = 0.13$  s,  $\alpha_2 = 0.14$ , and  $\tau_2 = 2.50$  s with  $\alpha = 0.10$  in Eq. 44).

we observed. We therefore performed an additional set of experiments using well-controlled synthetic conditioning stimuli to manipulate different stimulus properties independently.

For example, to examine the role of the frequency content of the conditioning stimulus, we first generated a dynamic moving ripple stimulus (Eqs. 1–3 in SYNTHETIC SOUNDS in METHODS) and then manipulated its frequency content by up- or downshifting its spectral components (Fig. 7; see also Supplemental Fig. S9). We could thereby generate conditioning stimulus ensembles in which only a particular sound property of interest was different, leaving all other characteristics unchanged. Thus in Fig. 7*A* frequency was varied but parameters such as intensity were unchanged. Using this approach we examined the effect of varying the following acoustic properties in conditioning stimuli: intensity (amplitude), frequency, amplitude-modulation (AM), frequency-modulation (FM), and higher-order spectrotemporal structure (see *Stimulus design* in METHODS and Fig. 1*B*). We used both natural and synthetic sounds as probe stimuli in these experiments, but found no difference between them and so combined the results in the population analysis (Fig. 8).

When we varied either intensities or frequencies—i.e., lower-order sound properties—in conditioning stimuli, we observed context-dependent effects in 77 probes (81.1%) of 95 probes (tested in 31 cells) and in 73/110 (64.6%; 35 cells), respectively (Fig. 8*A*). The effects were as large and long-lasting as those induced when natural sounds were used as conditioning stimuli (Fig. 8, *B* and *C*).

We then examined the effects of AM and FM changes in conditioning stimuli using modulated harmonic tones (Eqs. 4

and 5) and also the changes in even higher-order acoustic properties such as complex interactions between spectrotemporal sound elements by comparing the differences between modulated colored noise and its corresponding natural sounds (Eqs. 6–9). Context-dependent effects were observed in 62/96 (64.6%; 27 cells) for AM modulation; in 47/82 (57.3%; 25 cells) for FM modulation; and in 59/137 (43.1%; 27 cells) for colored noise modulation; but the effects were substantially smaller and shorter than the effects induced by the changes in natural sound contexts (Fig. 8*C*). That is, higher-order sound properties contributed to the context dependence mainly on a very short timescale, on the order of about 100 ms (Fig. 8*B*). From these population results, we conclude that neural responses in area A1 are more sensitive to changes in lower-order sound properties such as overall intensities and frequencies than to changes in higher-order properties such as amplitude- and frequency-modulations.

#### Subcortical contribution to context dependence

We have concluded that stimulus context can exert significant effects on the timescale of seconds in area A1 (see *Context dependence* and *Context dependence of synaptic conductances*). This context dependence could originate in the cortex (as suggested by previous work; Creutzfeldt et al. 1980; Miller et al. 2002; Ulanovsky et al. 2003, 2004; Wehr and Zador 2005) or could be inherited from thalamic response properties.

To test this in our preparation, we used the loose cell-attached patch method to record extracellularly from well-isolated single units in the auditory thalamus (MGB). Because firing rates in MGB were typically high (spontaneous,  $0.78 \pm 1.25$  Hz; evoked,  $11.4 \pm 16.9$  Hz; mean  $\pm$  SD), here we could obtain good estimates of stimulus-evoked activity from the average firing rate, without examining the subthreshold responses.

Figure 9*A* shows a typical example of a thalamic unit in response to a sequence of two natural sound fragments. As observed in the subthreshold responses in area A1 (Figs. 2 and 3), changing the preceding conditioning stimulus caused a difference in the suprathreshold responses to the following probe stimulus. However, in this example the effect of the conditioning stimulus was limited to the first bin (100 ms) of the poststimulus time histograms (PSTHs in Fig. 9*B*; see also Supplemental Fig. S10), indicating a rapid decay of the context-dependence effect.

To quantify the effect over the population, we first computed the SD of the PSTHs to each probe stimulus over all different conditioning stimuli (Fig. 9*C*) and then computed the average over all probes examined across the population of thalamic units (Fig. 9*D*; see also Eq. 42 in CONTEXT DEPENDENCE AT SUPRATHRESHOLD LEVEL). We found that the neuronal responses in MGB depended only on a short timescale ( $\tau = 80$  ms; 93 probes tested in 14 cells). Therefore we conclude that the contribution of subcortical adaptation to the cortical effects reported earlier is minimal and that the long-lasting component arises mainly in the cortex.

#### DISCUSSION

We have used in vivo whole cell patch-clamp recordings to study how stimulus history affects neural responses—and thus

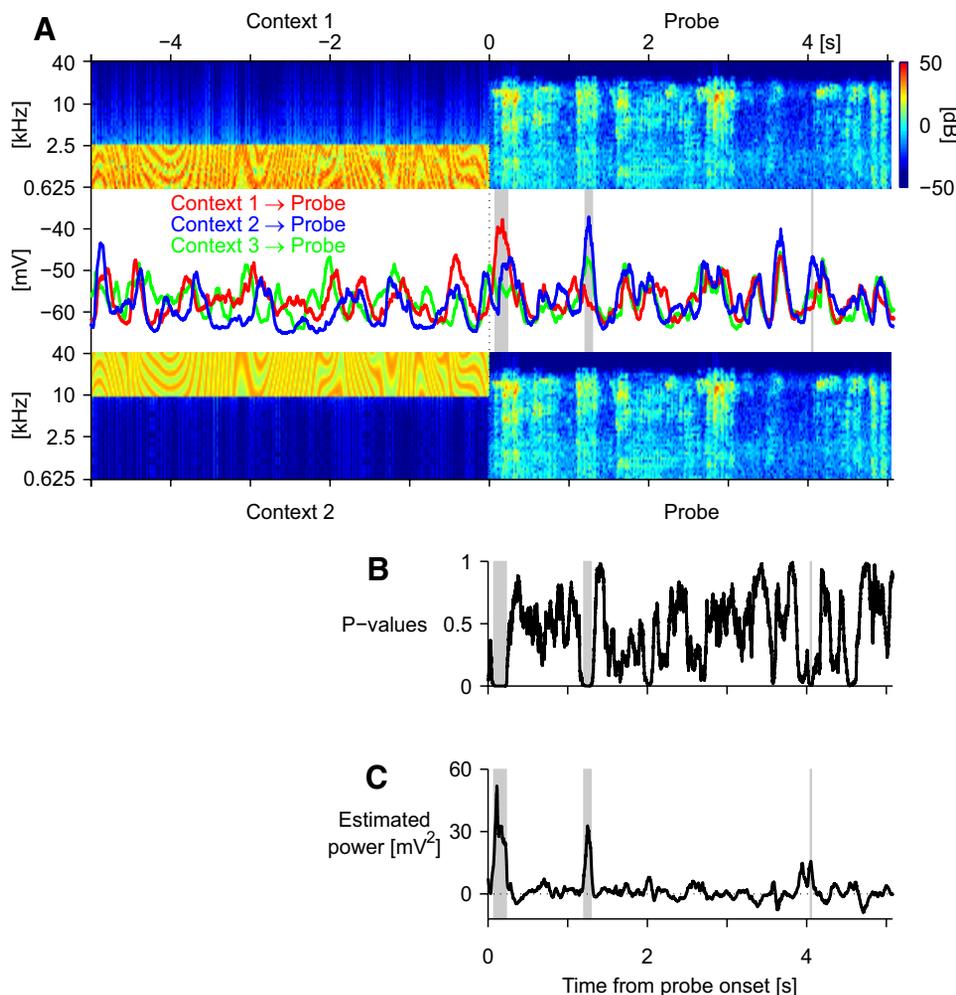


FIG. 7. Context dependence induced by frequency changes. *A*: typical mean subthreshold responses of a rat A1 neuron to part of a sound sequence, in which a natural sound stimulus probe was preceded by 3 different synthetic conditioning stimuli with different bandwidths [red, blue, and green lines for contexts with 0.625–2.5 kHz (*top spectrogram*), 10–40 kHz (*bottom spectrogram*), and 2.5–10 kHz (see Supplemental Fig. S9), respectively; *time 0* indicates the transition from conditioning to probe stimuli]. Spikes were clipped by a median filter (window length, 10 ms). Significant context dependence was observed for >4 s (gray bands; see *B* for details). *B*: black line represents *P* values and gray bands show the time points where the context dependence was statistically significant under the criterion:  $P < 0.01$  for  $\geq 5$  ms (in the same format as Fig. 2*D*). *C*: the response power estimate that depends on the context (black line; in the same format as that in Fig. 2*E*). The population average of this quantity normalized by the average predictable power is shown in Fig. 8*B* (the 2nd panel from the top).

constrains neuronal encoding models—in the primary auditory cortex (area A1). We found that temporal context can exert rather long-lasting effects—sometimes as long as 4 s. Even though sound-evoked synaptic conductances rarely lasted longer than around 100 ms, both excitatory and inhibitory components of synaptic inputs to A1 neurons showed long-lasting context dependence, suggesting that presynaptic mechanisms such as synaptic depression and/or facilitation would be involved. Restricting knowledge of the stimulus history to only a few hundred milliseconds reduced the predictable component of the response by about half. However, extending the time horizon did not lead to an appreciable increase in the performance of linear STRF-based models, indicating that the long-lasting effects of context were nonlinear. Thalamic recordings revealed that this long-lasting context dependence originated in the cortex. Our results demonstrate the importance of long-range temporal effects in auditory cortex and suggest a potential neural substrate for stream segregation and other forms of auditory processing that require integration over timescales of seconds or longer.

#### Context dependence and model construction

A central aim of this study was to characterize the length of the “memory” of auditory cortical neurons. To achieve this goal, we developed a novel experimental approach that al-

lowed us to quantify the importance of long-lasting contextual effects within the framework of input–output model construction.

Our approach differs from previous studies in at least three significant ways. First, we assessed the effect of context using spectrotemporally complex stimuli, rather than simpler stimuli such as pure or AM/FM tones and clicks as in many previous studies (Abeles and Goldstein 1972; Bartlett and Wang 2005; Brosch and Schreiner 1997, 2000; Calford and Semple 1995; Hocherman and Gilat 1981; Phillips 1985; Pienkowski and Eggermont 2009; Ulanovsky et al. 2003, 2004; Wehr and Zador 2005; but see David et al. 2009). To do this we designed an efficient stimulus protocol in which each sound serves double duty, both as a probe for the previous context and as a context for the following probe (see *Stimulus design* in METHODS). We could therefore directly address the role of temporal context for determining responses to arbitrary natural stimuli and quantified its effects from the viewpoint of response prediction (rather than examining the parameter changes in linear STRF-based models; David et al. 2009; Pienkowski and Eggermont 2009). Second, we monitored subthreshold rather than suprathreshold responses. Because neurons in area A1 are highly selective, firing rates to most stimuli are typically low under our experimental conditions (see also Hromádka et al. 2008; Wehr and Zador 2005). By using the subthreshold responses, however, we were able to generate reliable esti-

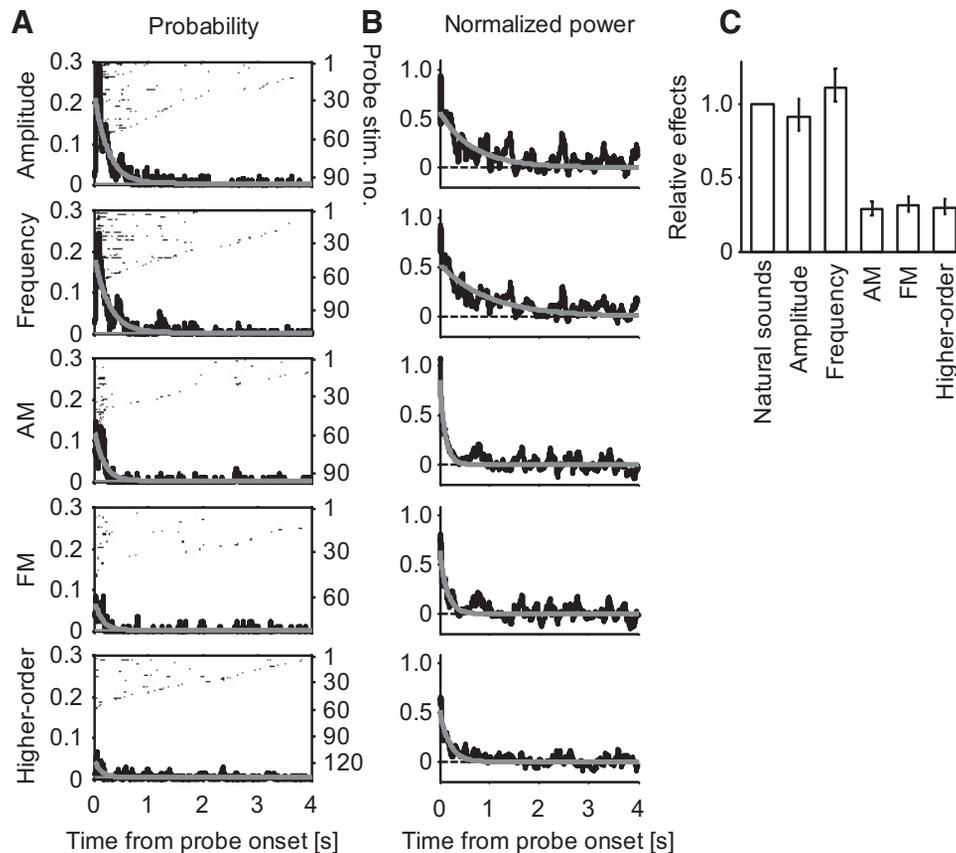


FIG. 8. Relation between context dependence and sound properties. Using synthetic sounds, we examined the effects of the changes in a particular sound property of interest on the responses to following probe stimuli. *Top to bottom*: the population analysis for the changes in amplitude (95 probes in 31 cells), frequency (110 probes in 35 cells), AM rates and depths (96 probes in 27 cells), FM rates and depths (82 probes in 25 cells), and higher-order properties (137 probes in 27 cells). The thick gray curves show the exponential curve fit as in Eq. 44, with the constant  $\alpha$  equal to the mean noise floor over time for A (thin gray) and with  $\alpha = 0$  for B (broken). *A*: significance measure. Shown is the population analysis based on the probability of observing context dependence, in the same format as that in Fig. 3A, but overlaid. *Top to bottom*: the parameters for the exponential model were  $[\alpha_1, \tau_1, \alpha] = [0.23, 0.25, 2.8 \times 10^{-3}]$ ,  $[0.19, 0.27, 2.3 \times 10^{-3}]$ ,  $[0.13, 0.17, 2.3 \times 10^{-3}]$ ,  $[0.07, 0.15, 2.1 \times 10^{-3}]$ , and  $[0.04, 0.13, 3.6 \times 10^{-3}]$ , respectively. *B*: fractional power measure. Shown in the population analysis, indicating the contribution of context dependence to the response dynamics, in the same format as that in Fig. 3B. *Top to bottom*: the parameters for the exponential curves (all with  $\alpha = 0$ ) were  $[\alpha_1, \tau_1] = [0.57, 0.66]$ ,  $[0.54, 0.97]$ ,  $[0.86, 0.11]$ ,  $[0.64, 0.15]$ , and  $[0.52, 0.20]$ , respectively. *C*: relative context-dependence effects. We computed the area under the curve (Eq. 21; from 0 to 4 s) of the normalized context-dependent power examined with various synthetic stimulus ensembles (black lines in B) and normalized it by the corresponding area examined with natural sound ensembles (Fig. 3B, bottom). For details, see CONTEXT DEPENDENCE AT SUBTHRESHOLD VOLTAGE LEVEL IN METHODS. The area itself represents well the total effects induced by ignoring the conditioning stimuli and the normalized area shows the relative context-dependence effects for each of the following stimulus properties (from left to right, with the 95% confidence intervals computed by resampling methods): natural sounds (bar height of one), amplitude, frequency, AM, FM, and higher-order acoustic property.

mates of the response from only a few presentations of each stimulus. Third, we decomposed synaptic inputs into excitatory and inhibitory components and analyzed context-dependence effects separately (see also Monier et al. 2008; Wehr and Zador 2003, 2005). In this way we could better speculate the mechanisms underlying the long-lasting context-dependence effects in area A1 (see *Mechanisms*).

Our analysis revealed a rather long window ( $\tau = 1.04$  s; Fig. 4) over which temporal context exerts its effect in A1. Building on studies in area A1 demonstrating that forward suppression and facilitation decay within a few hundred milliseconds (e.g., Brosch and Schreiner 1997, 2000; Calford and Semple 1995; but see Bartlett and Wang 2005; Pienkowski and Eggermont 2009; Ulanovsky et al. 2003, 2004; Wehr and Zador 2005), most linear encoding models typically use a window that is only a few hundred milliseconds long. Our results demonstrate that for spectrotemporally rich stimulus ensembles this is a period so short that even the best nonlinear model could not

hope to capture more than about half of the predictable component of the subthreshold response. Since linear models have largely failed to predict responses to spectrotemporally complex stimuli (Ahrens et al. 2008; Machens et al. 2004; Sahani and Linden 2003), we expected that extending the length of the stimulus history available to the linear model would improve performance.

We found, however, that incorporating a longer time horizon into the model yielded only a modest improvement in model performance. Figure 4 provides a detailed accounting of the various sources of model error. About half of the response power is predictable from even brief (<100-ms) segments of the stimulus. However, less than half of that (i.e., <20% of the total) is accessible to the optimal linear model and only slightly more to a linear model with a static nonlinearity. Our results thus suggest that STRF-based models are limited not only by the length of the stimulus history, but also by their simplicity, i.e., by their linearity (see also Ahrens et al. 2008).

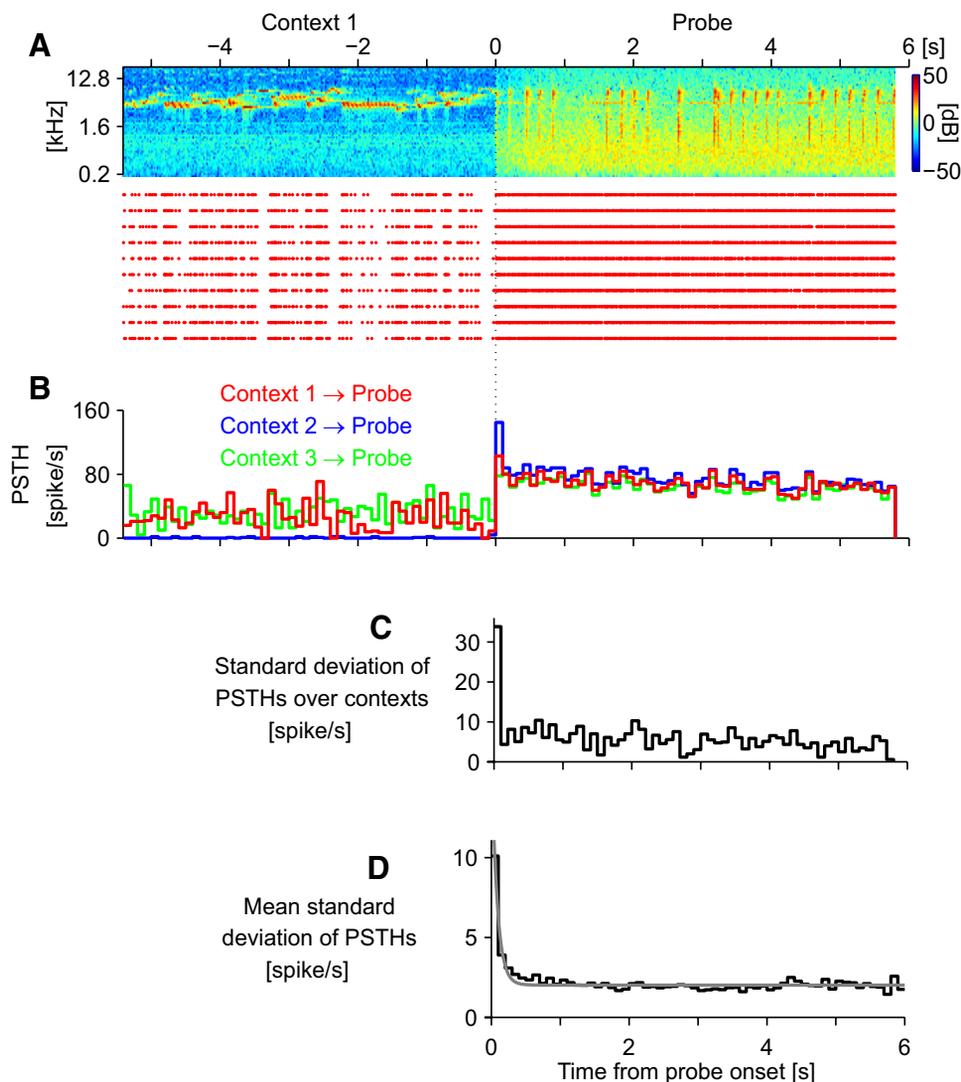


FIG. 9. Rapid decay of context dependence in auditory thalamus. *A*: typical suprathreshold responses of a neuron in rat auditory thalamus to part of a natural sound sequence (spectrogram; the same example in Fig. 2) over 10 repeats. Rasters in red indicate spike occurrences on individual trials. *B*: the poststimulus time histogram (PSTH); red line for the one shown in *A*, blue line for the one in response to the same probe stimulus following “silence” period, and green line for the one in response to the same probe stimulus but preceded by a different conditioning stimulus (see also Supplemental Fig. S10). Substantially different responses were observed only in the first bin after the onset of the probe stimulus ( $0 < t \leq 100$  ms). *C*: we assessed the context dependence at the spike level by computing the SD of PSTHs to the probe stimulus over all different contexts (black). *D*: average SD of the PSTHs—as in *C*—over the population (93 probes in 14 cells, black; Eq. 42). Context dependence in auditory thalamus lasted only a short period of time. The decay size and constant are:  $\alpha_1 = 15.0$ ,  $\tau_1 = 80$  ms with  $\alpha = 2.0$  spike/s (gray; Eq. 44).

What kinds of nonlinearities might be needed? At one extreme, the nonlinearities might be involved in adaptation; adaptation on various timescales is ubiquitous in sensory systems (Baccus and Meister 2002; Carandini and Ferster 1997; Dean et al. 2005; Fairhall et al. 2001; Kvale and Schreiner 2004; Movshon and Lennie 1979; Müller et al. 1999; Ulanovsky et al. 2004). Such adaptive nonlinearities might be relatively simple in form. At the other extreme, A1 neurons might detect acoustic “features” such as edges or even more complex high-order acoustic invariants (Fishbach et al. 2001, 2003). An intermediate possibility is that A1 neurons might implement both kinds of nonlinearities, but at different timescales: over relatively short time periods (e.g.,  $< 100$  ms) they might act as feature detectors, whereas on longer timescales simpler forms of adaptation operate. In support of this view is the time course of the adaptation in Fig. 4: a very fast ( $< 100$ -ms) time constant that accounts for 50% of the predictable response and a slower one for the remainder.

### Mechanisms

What mechanisms might underlie the long-lasting history dependence we observed? Our data provide at least three clues

to address this question. First, consistent with previous results (Creutzfeldt et al. 1980; Miller et al. 2002; Ulanovsky et al. 2003, 2004; Wehr and Zador 2005), the long-lasting effects were absent in the auditory thalamus (Fig. 9) and thus unlikely to be inherited from thalamic inputs. Second, the decay constant we observed in A1 was on the order of seconds for both excitatory and inhibitory components of the responses (Fig. 6) and it was much longer than the membrane constant of neurons or the duration of sound-evoked synaptic events ( $\sim 100$  ms; but see Carandini and Ferster 1997; Sanchez-Vives et al. 2000). Third, learning was unlikely to be involved under our experimental conditions (see, in contrast, Fritz et al. 2003, 2005). We thus think it is likely that native cortical network properties, acting via synaptic depression and/or facilitation (Abbott et al. 1997; David et al. 2009; Tsodyks and Markram 1997; Wehr and Zador 2005; Zucker and Regehr 2002), were largely responsible for the long-lasting effects of stimulus history and its context in A1.

### Functional implications

What functional role might such long-lasting context dependence play? One important role of adaptation is to increase the

effective dynamic range of a sensory neuron (Brenner et al. 2000; Fairhall et al. 2001). Sensory neurons typically have a dynamic range greatly exceeded by that of the sensory environment. The timescale of adaptation typically depends on stimulus statistics and the direction of the changes (Baccus and Meister 2002; David et al. 2009; DeWeese and Zador 1998; Kvale and Schreiner 2004; Pienkowski and Eggermont 2009; Scholl et al. 2008; Smirnakis et al. 1997; Wehr and Zador 2005). In fact we found that differences in stimulus intensity and bandwidth had stronger and longer effects on the neuronal responses to the following stimuli than those in AM, FM, or other higher-order spectrotemporal modulations (Fig. 8). Adaptation can thus provide a means of making efficient use of limited sensory bandwidth.

The context dependence we describe is most closely related to stimulus-specific adaptation (SSA), which can be observed in a paradigm in which a rare (“oddball”) probe stimulus is intermixed with a more common conditioning stimulus. Neurons in A1 respond to the probe stimulus more strongly when it is rarer, consistent with the proposal that SSA is a mechanism for enhancing rare foreground events from a more homogeneous acoustic background (Ulanovsky et al. 2003, 2004). The context dependence we describe differs from SSA in that our experimental design includes no explicit common or rare stimuli—it is elicited by a very broad range of complex stimuli—and so cannot be readily interpreted in terms of foreground and background. Nevertheless, the context dependence we describe may represent a generalization of SSA and they may share similar or even identical mechanisms.

From a functional viewpoint, detecting an oddball stimulus in a noisy background may represent a specialized computation required to perform stream segregation (Bregman 1990). Psychophysical experiments indicate that sensory memory of sounds—or “echoic memory” (Neisser 1967)—typically persists on the order of a few seconds (Glucksberg and Cowen 1970; Kubovy and Howard 1976). It is then tempting to speculate that both SSA and context-dependent adaptation represent neural correlates of stream segregation. Furthermore, the integration of stimulus history and its context in area A1 might contribute to many other auditory perceptual tasks, including speech processing, pitch/rhythm detection, and music expectation. They all require extracting certain spectrotemporal patterns in acoustic stimuli over seconds and, importantly, they all depend on context. Because the responses in area A1 are in general highly selective (Hromádka et al. 2008), however, the processing in A1 per se would be insufficient to fully perform such perceptual tasks. Nevertheless, the presence of long time constants in A1 but not in auditory thalamus suggests that area A1 might play a critical role in auditory perception by forming building blocks for processing at later stages.

#### ACKNOWLEDGMENTS

We thank M. Wehr for providing the recording data for the spectrotemporal receptive field-based model analysis, H. Oviedo and T. Hromádka for extensive help in experiments, and all the members of the Zador laboratory for many useful discussions and comments on the manuscript.

Present address of H. Asari: Harvard University, 52 Oxford Street, Cambridge, MA 02138.

#### GRANTS

This work was supported by a Farish-Gerry Fellowship to H. Asari and grants from the National Institutes of Health, the Swartz Foundation, the Marie Robertson Fund, and the Morin Trust to A. M. Zador.

#### REFERENCES

- Abbott LF, Varela JA, Sen K, Nelson SB.** Synaptic depression and cortical gain control. *Science* 275: 220–224, 1997.
- Abeles M, Goldstein MH.** Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. *Brain Res* 42: 337–352, 1972.
- Ahrens MB, Linden JF, Sahani M.** Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci* 28: 1929–1942, 2008.
- Baccus SA, Meister M.** Fast and slow contrast adaptation in retinal circuitry. *Neuron* 36: 909–919, 2002.
- Barry PH.** JPCalc, a software package for calculating liquid junction potential corrections in patch-clamp, intracellular, epithelial and bilayer measurements and for correcting junction potential measurements. *J Neurosci Methods* 51: 107–116, 1994.
- Bartlett EL, Wang X.** Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 94: 83–104, 2005.
- Bar-Yosef O, Rotman Y, Nelken I.** Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J Neurosci* 22: 8619–8632, 2002.
- Bregman AS.** *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: MIT Press, 1994.
- Brenner N, Bialek W, de Ruyter van Steveninck R.** Adaptive rescaling maximizes information transmission. *Neuron* 26: 695–702, 2000.
- Brosch M, Schreiner C.** Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 77: 923–943, 1997.
- Brosch M, Schreiner C.** Sequence sensitivity of neurons in cat primary auditory cortex. *Cereb Cortex* 10: 1155–1167, 2000.
- Calford MB, Semple MN.** Monaural inhibition in cat auditory cortex. *J Neurophysiol* 73: 1876–1891, 1995.
- Carandini M, Ferster D.** A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* 276: 949–952, 1997.
- Carnevale NT, Johnston D.** Electrophysiological characterization of remote chemical synapses. *J Neurophysiol* 47: 606–621, 1982.
- Chichilnisky EJ.** A simple white noise analysis of neuronal light responses. *Network* 12: 199–213, 2001.
- Cleveland WS.** Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 74: 829–836, 1979.
- Cleveland WS, Devlin SJ.** Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 83: 596–610, 1988.
- Cohen L.** *Time-Frequency Analysis.* Englewood Cliffs, NJ: Prentice Hall, 1995.
- Creutzfeldt O, Hellweg FC, Schreiner C.** Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res* 39: 87–104, 1980.
- David SV, Mesgarani N, Fritz JB, Shamma SA.** Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *J Neurosci* 29: 3374–3386, 2009.
- Dean I, Harper NS, McAlpine D.** Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8: 1684–1689, 2005.
- Depireux DA, Simon JZ, Klein DJ, Shamma SA.** Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220–1234, 2001.
- DeWeese MR, Zador AM.** Asymmetric dynamics in optimal variance adaptation. *Neural Comput* 10: 1179–1202, 1998.
- Eggermont JJ, Johannesma PM, Aertsen AM.** Reverse-correlation methods in auditory research. *Q Rev Biophys* 16: 341–414, 1983.
- Escabi MA, Schreiner CE.** Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22: 4114–4131, 2002.
- Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR.** Efficiency and ambiguity in an adaptive neural code. *Nature* 412: 787–792, 2001.
- Fishbach A, Nelken I, Yeshurun Y.** Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol* 85: 2303–2323, 2001.
- Fishbach A, Yeshurun Y, Nelken I.** Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. *J Neurophysiol* 90: 3663–3678, 2003.

- Fritz J, Elhilali M, Shamma S.** Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res* 206: 159–176, 2005.
- Fritz J, Shamma S, Elhilali M, Klein D.** Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6: 1216–1223, 2003.
- Garcia-Lazaro JA, Ahmed B, Schnupp JWH.** Tuning to natural stimulus dynamics in primary auditory cortex. *Curr Biol* 16: 264–271, 2006.
- Gill P, Zhang J, Woolley SMN, Fremouw T, Theunissen FE.** Sound representation methods for spectro-temporal receptive field estimation. *J Comput Neurosci* 21: 5–20, 2006.
- Glucksberg S, Cowen GN.** Memory for nonattended auditory material. *Cogn Psychol* 1: 149–156, 1970.
- Hastie T, Tibshirani R, Friedman J.** *The Elements of Statistical Learning Theory*. New York: Springer, 2001.
- Hocherman S, Gilat E.** Dependence of auditory cortex evoked unit activity on interstimulus interval in the cat. *J Neurophysiol* 45: 987–997, 1981.
- Hromádka T, DeWeese MR, Zador AM.** Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol* 6: e16, 2008.
- Joris PX, Schreiner CE, Rees A.** Neural processing of amplitude-modulated sounds. *Physiol Rev* 84: 541–577, 2004.
- Klein DJ, Depireux DA, Simon JZ, Shamma SA.** Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9: 85–111, 2000.
- Koch C, Poggio T, Torre V.** Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philos Trans R Soc Lond B Biol Sci* 298: 227–263, 1982.
- Kowalski N, Depireux DA, Shamma SA.** Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *J Neurophysiol* 76: 3524–3534, 1996.
- Kruskal WH, Wallis WA.** Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 47: 583–621, 1952.
- Kubovy M, Howard FP.** Persistence of pitch-segregating echoic memory. *J Exp Psychol* 2: 531–537, 1976.
- Kvale MN, Schreiner CE.** Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol* 91: 604–612, 2004.
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM.** Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol* 90: 2660–2675, 2003.
- Machens CK, Wehr MS, Zador AM.** Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24: 1089–1100, 2004.
- Miller LM, Escabí MA, Read HL, Schreiner CE.** Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87: 516–527, 2002.
- Monier C, Fournier J, Frégnac Y.** *In vitro* and *in vivo* measures of evoked excitatory and inhibitory conductance dynamics in sensory cortices. *J Neurosci Methods* 169: 323–365, 2008.
- Movshon JA, Lennie P.** Pattern-selective adaptation in visual cortical neurons. *Nature* 278: 850–852, 1979.
- Müller JR, Metha AB, Krauskopf J, Lennie P.** Rapid adaptation in visual cortex to the structure of images. *Science* 285: 1405–1408, 1999.
- Neisser U.** *Cognitive Psychology*. New York: Meredith, 1967.
- Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D.** Primary auditory cortex of cats: feature detection or something else? *Biol Cybern* 89: 397–406, 2003.
- Paninski L.** Convergence properties of three spike-triggered analysis techniques. *Network* 14: 437–464, 2003.
- Phillips DP.** Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. *Hear Res* 19: 253–268, 1985.
- Pienkowski M, Eggermont J.** Effects of adaptation on spectrotemporal receptive fields in primary auditory cortex. *NeuroReport* 20: 1198–1203, 2009.
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP.** Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46: 945–956, 2005.
- Sahani M, Linden JF.** How linear are auditory cortical responses? In: *Advances in Neural Information Processing Systems*, edited by Becker S, Thrun S, Obermeyer K. Cambridge, MA: MIT Press, 2003, vol. 15, p. 109–116.
- Sanchez-Vives MV, Nowak LG, McCormick DA.** Cellular mechanisms of long-lasting adaptation in visual cortical neurons *in vitro*. *J Neurosci* 20: 4286–4299, 2000.
- Scholl B, Gao X, Wehr M.** Level dependence of contextual modulation in auditory cortex. *J Neurophysiol* 99: 1616–1627, 2008.
- Schwartz O, Simoncelli EP.** Natural signal statistics and sensory gain control. *Nat Neurosci* 4: 819–825, 2001.
- Sharpee T, Rust NC, Bialek W.** Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16: 223–250, 2004.
- Simoncelli EP, Pillow J, Paninski L, Schwartz O.** Characterization of neural responses with stochastic stimuli. In: *The Cognitive Neurosciences III*, edited by Gazzaniga M. Cambridge, MA: MIT Press, 2004, chapt. 23, p. 327–338.
- Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M.** Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386: 69–73, 1997.
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL.** Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12: 289–316, 2001.
- Tsodyks MV, Markram H.** The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci USA* 94: 719–723, 1997.
- Ulanovsky N, Las L, Farkas D, Nelken I.** Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24: 10440–10453, 2004.
- Ulanovsky N, Las L, Nelken I.** Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6: 391–398, 2003.
- van Lint JH.** *Introduction to Coding Theory*. New York: Springer-Verlag, 1992.
- Wehr M, Zador AM.** Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426: 442–446, 2003.
- Wehr M, Zador AM.** Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron* 47: 437–445, 2005.
- Wu MCK, David SV, Gallant JL.** Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505, 2006.
- Zador AM, Agmon-Snir H, Segev I.** The morphoelectronic transform: a graphical approach to dendritic function. *J Neurosci* 15: 1669–1682, 1995.
- Zucker RS, Regehr WG.** Short-term synaptic plasticity. *Annu Rev Physiol* 64: 355–405, 2002.