

Inferring the role of inhibition in auditory processing of complex natural stimuli

Nadja Schinkel-Bielefeld,¹ Stephen V. David,² Shihab A. Shamma,² and Daniel A. Butts¹

¹Department of Biology and Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland; and ²Institute for Systems Research, University of Maryland, College Park, Maryland

Submitted 21 December 2011; accepted in final form 23 March 2012

Schinkel-Bielefeld N, David SV, Shamma SA, Butts DA. Inferring the role of inhibition in auditory processing of complex natural stimuli. *J Neurophysiol* 107: 3296–3307, 2012. First published March 28, 2012; doi:10.1152/jn.01173.2011.—Intracellular studies have revealed the importance of cotuned excitatory and inhibitory inputs to neurons in auditory cortex, but typical spectrotemporal receptive field models of neuronal processing cannot account for this overlapping tuning. Here, we apply a new nonlinear modeling framework to extracellular data recorded from primary auditory cortex (A1) that enables us to explore how the interplay of excitation and inhibition contributes to the processing of complex natural sounds. The resulting description produces more accurate predictions of observed spike trains than the linear spectrotemporal model, and the properties of excitation and inhibition inferred by the model are furthermore consistent with previous intracellular observations. It can also describe several nonlinear properties of A1 that are not captured by linear models, including intensity tuning and selectivity to sound onsets and offsets. These results thus offer a broader picture of the computational role of excitation and inhibition in A1 and support the hypothesis that their interactions play an important role in the processing of natural auditory stimuli.

model; nonlinear; audition; cortex

AS THE STANDARD FIRST-ORDER description of the stimulus selectivity of an auditory neuron, the linear spectrotemporal receptive field (STRF) offers an intuition for how the neuron responds to complex stimuli. Sounds that most closely match the STRF (in both frequency and relative timing) will drive the strongest responses, whereas sounds that do not match the spectrotemporal pattern of the STRF will drive the neuron weakly or not at all (Aertsen and Johannesma 1981; deCharms et al. 1998; Kowalski et al. 1996). Thus the STRF can be used to understand the role a neuron plays in representing sounds as well as predict neuronal responses in a variety of contexts.

However, recent intracellular studies in auditory cortex challenge the applicability of such straightforward stimulus processing to primary auditory cortex (A1) neurons. These studies show that inhibitory inputs onto A1 neurons are often cotuned with excitation (Dorn et al. 2010; Sun et al. 2010; Wehr and Zador 2003) and implicate the interplay of excitation and inhibition in particular nonlinear aspects of A1 neuron responses such as precise timing (Wehr and Zador 2003) and intensity tuning (Wu et al. 2006). Cotuned excitation and inhibition cannot be directly derived from the linear STRF, which implicitly assumes that neural responses are derived from an average over the many inputs of the neuron. Such averaging necessarily hides the effects of excitation and inhibition where their tuning overlaps (Christianson et al. 2008).

Likewise, studies of A1 in complex stimulus contexts have revealed that responses of A1 neurons can be highly nonlinear (David et al. 2009; Machens et al. 2004; Sahani and Linden 2003). Nonlinear tuning may be particularly important in shaping responses to natural stimuli, as these have rich spectrotemporal dynamics. Recent modeling studies have generally addressed some nonlinear aspects of responses in A1 (Ahrens et al. 2008a; Atencio et al. 2008; David et al. 2009), but none has directly addressed how the interplay of excitation and inhibition might play a role in its nonlinear processing.

Here, we employ a novel nonlinear modeling approach designed to infer separate tuning and response properties of excitation and inhibition from extracellular recordings of neurons in A1 during the presentation of speech stimuli (David et al. 2009). We find that the inferred tuning of excitation and inhibition can predict a variety of neuronal response characteristics that have been identified in A1, including sharp transient responses to sound onset and offset and nonmonotonic intensity tuning (Greenwood and Maruyama 1965; Phillips et al. 1995; Schreiner et al. 1992). By capturing both the spectrotemporal tuning and nonlinear interactions between putative excitatory and inhibitory elements, the resulting model combines several important aspects of auditory processing in a unified computational framework.

MATERIALS AND METHODS

Experimental procedures. Auditory responses were recorded extracellularly from single neurons in A1 of six awake, passively listening ferrets. As described in detail in David et al. (2009), animals were implanted with a steel headpost to allow for stable recording. While under anesthesia (ketamine and isoflurane), the skin and muscles on the top of the head were retracted from the central 4-cm diameter of the skull. Several titanium bone screws were attached to the skull, a custom metal post was glued on the midline, and the entire site was covered with bone cement. After surgery, the skin around the implant was allowed to heal. After recovery, a small craniotomy (1–2 mm diameter) was opened through the cement and skull over auditory cortex. The protocol for all surgical and experimental procedures was approved by the institutional animal care and use committee at the University of Maryland and is consistent with National Institutes of Health guidelines.

Single-unit activity was recorded using tungsten microelectrodes (1–5 M Ω ; FHC) from head-fixed animals in a double-walled, sound-attenuating chamber (Industrial Acoustics). Data were recorded simultaneously from one to four electrodes controlled by independent microdrives using a commercial data acquisition system (Alpha-Omega). The raw electrophysiological trace was digitized and band-pass-filtered between 300 and 6,000 Hz. Spiking events were extracted from the continuous signal using principal components analysis and *k*-means clustering. Only clusters with 90% isolation (i.e., 90% spikes in the cluster were likely to have been produced by a single neuron) were used for analysis. After identification of a recording site with isolatable units, a sequence of random tones (100-ms

Address for reprint requests and other correspondence: D. A. Butts, Dept. of Biology, 1210 Biology-Psychology Bldg. #144, Univ. of Maryland, College Park, MD 20742 (e-mail: dab@umd.edu).

duration, 500-ms separation) was used to measure latency, spectral tuning, and in some cases intensity tuning. Neurons were verified as being in A1 by their tonotopic organization, latency, and simple frequency tuning (Bizley et al. 2005).

Stimuli. We recorded the responses of isolated A1 neurons to segments of continuous speech. Speech samples were taken from the Texas Instruments/Massachusetts Institute of Technology (TIMIT) database (Garofolo 1988). Stimuli were sentences (3–4 s) sampled each from a different speaker and balanced across sex. The digital signals were transformed to analog, equalized to achieve flat gain, and amplified to a calibrated level. Stimuli were presented through an earphone (Etymotics) contralateral to the recording site. Before each experiment, the equalizer was calibrated to achieve flat gain according to the acoustic properties of the earphone insertion. All stimuli were presented at 65-dB peak sound pressure level (SPL). The original stimuli were recorded at 16 kHz and upsampled to 40 kHz using linear interpolation before presentation. Note that, as a result, this means that all stimulus power was below 8 kHz.

In this study, we recorded from 2 sets of neurons. The 1st set was exposed to a shorter stimulus consisting of 30 sentences, each 3 s long. The 2nd set of neurons was presented with a longer stimulus that, in addition to the 1st stimulus, contained 30 additional sentences, each 4 s in duration. All stimuli were repeated 5 times for nearly all neurons, although in a few cases there were 4 or 6 repetitions. We performed all analyses using 5-ms resolution.

Maximum likelihood estimation. We used maximum likelihood estimation to calculate the linear STRFs of A1 neurons. The likelihood is the probability of the model given the observed data, and thus maximizing the likelihood (which also maximizes the logarithm of the likelihood) is equivalent to finding the model parameters that best explain the data. The standard log-likelihood is given by (Snyder and Miller 1991):

$$LL = \sum_{t_s} \log_2 r(t_s) - \sum_t r(t) \tag{1}$$

where $r(t)$ is the time-varying firing rate predicted by the model, and t_s is the set of observed spike times. Model parameters are optimized to predict a large $r(t)$ at t_s (making the 1st term in Eq. 1 large) and low/0 firing rate everywhere else (making the 2nd term small). The values of LL reported are adjusted by a baseline LL, defined by the LL of a model that predicted a stimulus-independent mean firing. As a result, the LL is larger to the degree that it achieves a better explanation of the data than this null model, and it is bounded above by the single-spike information (Kouh and Sharpee 2009). It is reported in units of bits per spike.

The generalized linear model. The STRF is the basis for a first-order model of the relationship between the stimulus and the neural response. Traditionally, the STRF is calculated for auditory neurons using normalized reverse correlation (i.e., the spike-triggered average), which in this case is adjusted for the correlation structure of the stimulus used (Theunissen et al. 2001). Here, we use maximum likelihood estimation in the context of the generalized linear model (GLM) to estimate the STRF (Paninski 2004), which automatically accounts for stimulus correlations and can also impart more flexibility with regards to regularization (Calabrese et al. 2011). Furthermore, it also allows for direct comparisons with the generalized nonlinear model (GNM) described below, which is estimated with the same approach. Other linear methods such as normalized reverse correlation (David et al. 2007; Theunissen et al. 2001) and boosting (David et al. 2007; Zhang and Yu 2005) implicitly use different cost functions (i.e., minimizing mean-squared error instead of maximizing model likelihood) but yield results similar to these GLM methods (data not shown).

Although the GLM approach provides the flexibility to include simultaneously other linear processing elements such as spike refractoriness (Paninski 2004), here we simply use it in the context of the

common linear-nonlinear (LN) model (Chichilnisky 2001), which is a cascade model with a predicted firing rate given by:

$$r(t) = F[\mathbf{k} * \mathbf{s}(t)] = F[g(t)] \tag{2}$$

The embedded linear stage, $g(t) = \mathbf{k} * \mathbf{s}(t) = \sum_{\tau} \mathbf{s}(f, t - \tau)k(f, \tau)$, reflects how closely the stimulus $\mathbf{s}(t)$ matches the STRF, \mathbf{k} . Note, we use boldface here to denote vectors. The output of STRF processing, $g(t)$, is then translated into a firing rate through the spiking nonlinearity, $F[\cdot]$, given by the analytic function:

$$F[g] = \log[1 + \exp(g - \theta)] \tag{3}$$

where θ is the spike threshold. The form of this nonlinearity guarantees that the likelihood as a function of model parameters has only one unique local maximum (i.e., a global maximum) and thus can be efficiently optimized using gradient ascent to find a the overall maximum likelihood (Paninski 2004).

A shortcoming of linear models is that, although neurons often are driven by multiple inputs, linear models cannot have more than a single receptive field because processing by two linear receptive fields, \mathbf{k}_1 and \mathbf{k}_2 , is mathematically equivalent to that of a single receptive field given by their sum: $[\mathbf{s}(t) * \mathbf{k}_1] + [\mathbf{s}(t) * \mathbf{k}_2] = (\mathbf{k}_1 + \mathbf{k}_2) * \mathbf{s}(t) = \mathbf{k}_{\text{eff}} * \mathbf{s}(t)$.

The intensity-tuned LN model. Linear models assume that the firing rate increases linearly with the power at a given frequency (or log intensity depending on the representation of the stimulus). However, A1 neuron responses can be saturating with intensity or maximum at a particular intensity (and decreasing for higher intensities). One simple way to fit the property of nonlinear intensity tuning is to incorporate explicitly an “input nonlinearity” into a model (Ahrens et al. 2008a,b) that rescales the stimulus intensity and is applied to the stimulus before processing by the STRF. This input nonlinearity, $f_{\text{in}}(\cdot)$, can also be efficiently optimized in the context of the GLM framework by representing the input nonlinearity as a linear combination of overlapping tent-basis functions, $f_{\text{in}}(x) = \sum_n \alpha_n \xi_n(x)$ (Ahrens et al. 2008b). Here, we choose to use tent basis functions for ξ_n , which specify a linear interpolation of the function $f_{\text{in}}(\cdot)$ specified by the coefficients α_n (Ahrens et al. 2008a). Because the coefficients α_n operate linearly on the processed stimulus $\xi[\mathbf{s}(t, f)]$, they can also be optimized as linear parameters. Bilinear optimization is performed by alternating between optimizing \mathbf{k} holding α constant and vice versa until an optimum is reached (Ahrens et al. 2008b).

The generalized nonlinear model. The GNM framework is composed of nonlinear elements designed to simulate the effects of excitation and inhibition and also can be fit to data using maximum likelihood estimation (Butts et al. 2011). It is a cascade model where the predicted firing rate $r(t)$ is given by (Fig. 1A, right):

$$r(t) = F[\mathbf{W}_{\text{ex}}(t) * \mathbf{h}_{\text{ex}} + \mathbf{W}_{\text{in}}(t) * \mathbf{h}_{\text{in}}] \tag{4}$$

where $F[\cdot]$ is the same spiking nonlinearity as described above (Eq. 3), and the two terms summed in the brackets describe the excitatory and inhibitory elements. Each element is constructed to resemble a synaptic input, consisting of how the presynaptic neuron processes the stimulus [resulting in $\mathbf{W}_x(t)$] and a “postsynaptic current,” \mathbf{h}_x , which describes how the input of the presynaptic neuron contributes to the response of the A1 neuron. To limit the complexity of the model, in this case each element represents the pooled excitatory or inhibitory input, although it is in principle possible to add further terms in a more general formulation of the GNM.

Also in the spirit of this model being a next-order expansion, we assume that the excitatory and inhibitory inputs, $\mathbf{W}_x(t)$, are simple linear processes followed by a nonlinearity (i.e., an LN model):

$$\mathbf{W}_x(t) = f_x[\mathbf{s}(t) * \mathbf{k}_x] \tag{5}$$

where $\mathbf{s}(t) * \mathbf{k}_x$ describes the linear processing by the excitatory or inhibitory STRF \mathbf{k}_x , and the internal nonlinearity $f_x(\cdot)$ captures the nonlinear processing inherent in the spiking nonlinearity of the pre-

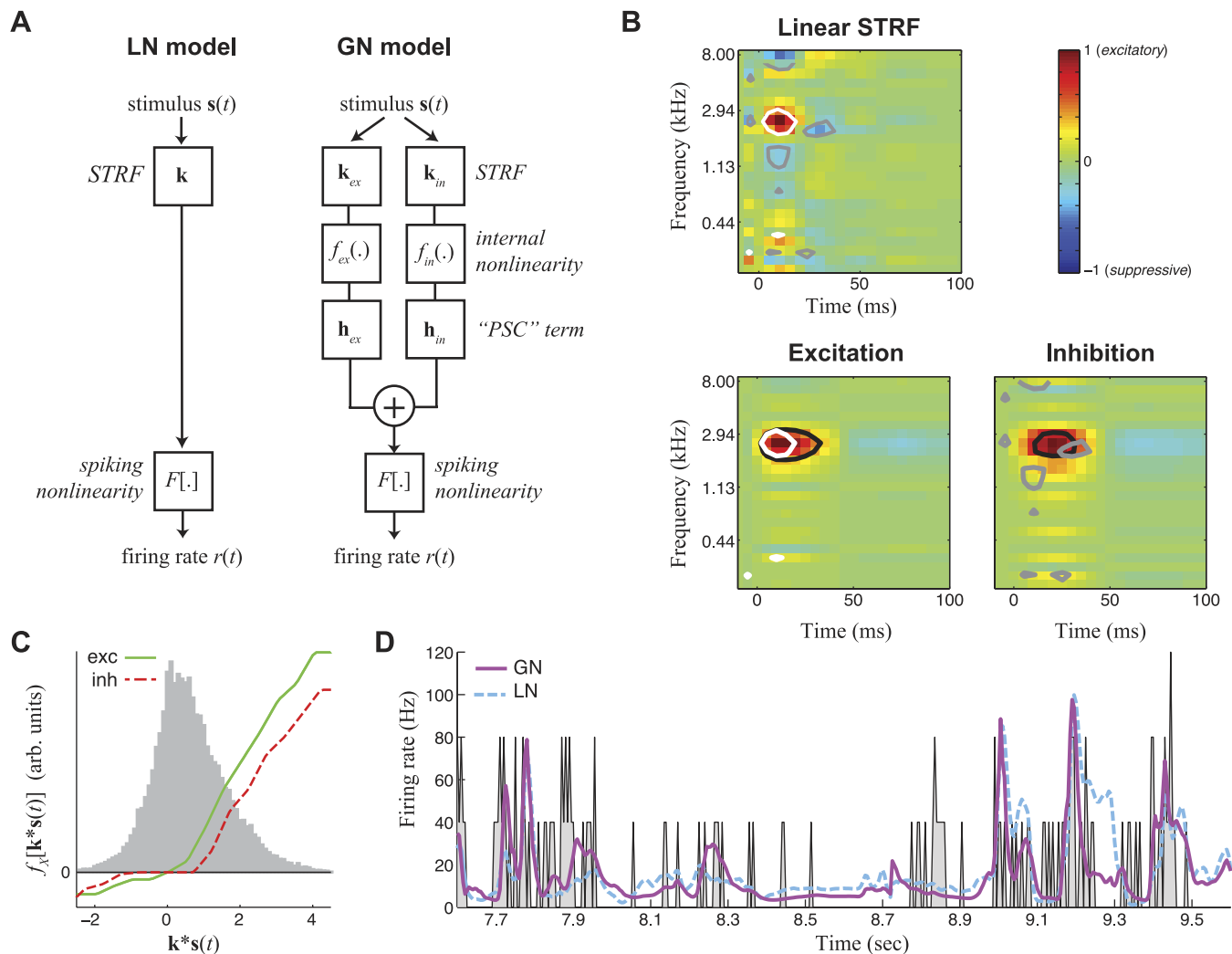


Fig. 1. Nonlinear modeling predicts overlapping excitation (ex) and inhibition (in). *A*: model schematics for the linear-nonlinear (LN) model compared with the generalized nonlinear (GN) model (GNM). Rather than attempting to capture all stimulus processing with a single linear spectrotemporal receptive field (STRF; *left*), the GNM (*right*) explicitly models excitatory and inhibitory inputs as separate LN elements, which are then linearly combined and passed through a spiking nonlinearity. PSC, postsynaptic current. *B*: the linear STRF (*top*) compared with separate STRFs for excitation and inhibition of the GNM (*bottom*). To compare the excitatory and inhibitory influences between the 2 models, white and gray contours show the 40% of the peak magnitude of excitatory and suppressive elements in the linear STRF (*top*) compared with black contours showing the 40% peak magnitude of the excitatory and inhibitory STRFs of the GNM (*bottom*). This illustrates that excitatory regions of the linear STRF (white) are smaller than the excitation predicted by the GNM (black, *bottom left*), and the linear STRF has small flanking suppressive regions (gray) compared with the broader inhibition in the GNM (black, *bottom right*) that largely overlaps with the GNM excitation. *C*: nonlinearities measured in the GNM for the excitatory (exc; green) and inhibitory (inh; red) module. The gray histogram shows the distribution of the filtered stimulus values for the training data, over which the lateral nonlinearities $f_x(\cdot)$ are applied. *D*: cross-validated peristimulus time histogram (PSTH) for the GNM (magenta) and the LN model (cyan). The measured firing rate is shown in gray.

vious cell (or group of cells). Note that the labels for each term are meant to describe the intuition behind the structure of the model, but we cannot claim that these terms necessarily represent actual excitatory and inhibitory currents (or conductances) based on this evidence alone because they might also be affected by other nonlinear aspects of neuronal processing that are not explicitly modeled.

As with the GLM, model parameters are chosen to maximize the LL of the model prediction given the data (Eq. 1). Furthermore, the structure of the GNM (Eq. 4) is such that the postsynaptic current terms \mathbf{h}_x are acting linearly on functions of the stimulus $\mathbf{W}_x(t)$ and thus can be efficiently estimated within the GLM framework described above. Similarly, the internal nonlinearity $f_x(\cdot)$ within $\mathbf{W}_x(t)$ can be cast as a linear combination of a set of nonlinear basis functions, $f_x(\cdot) = \sum_n \alpha_n \xi_n[\cdot]$, and their coefficients can also be efficiently optimized in the context of the GLM (Ahrens et al. 2008a,b). The postsynaptic current terms and internal nonlinearity coefficients constitute a bilinear model and are optimized by holding

one constant while fitting the second in an alternating fashion (Ahrens et al. 2008b). Linear constraints are introduced to enforce the nonlinearities $f_x(\cdot)$ to be monotonic and the postsynaptic current terms \mathbf{h}_x to be only positive (excitatory) or negative (inhibitory), which improved the fitting procedure and still resulted in the best final solutions. It also constrained the model parameters to the biologically plausible interpretation of excitatory and inhibitory inputs.

Thus, for a given choice of excitatory and inhibitory STRFs \mathbf{k}_x , all other model parameters (i.e., all \mathbf{h}_x and f_x) have uniquely specified optimal values through efficient GLM optimization, with a single LL. As a result, each choice of the excitatory and inhibitory STRFs has a LL associated with it, and they can thus be optimized through gradient ascent calculation as well (Butts et al. 2011).

Unlike the other model parameters, this optimization of the STRFs \mathbf{k}_x is not guaranteed to yield a global maximum of the LL. Thus it is important to start the fit with a reasonable guess for the receptive fields, which is constructed from the space-time separable STRF

derived with the GLM. The initial guess for both the excitatory and the inhibitory temporal kernel of the GNM is identical to the temporal kernel of the GLM. The initial guess for the frequency kernels differs for excitation and inhibition, with the positive parts $\mathbf{k}_f^{\text{GLM}|+}$ of the GLM-frequency kernel $\mathbf{k}_f^{\text{GLM}}$ used for the initial guess for the excitatory kernel in the GNM. The initial inhibitory kernel in the GNM is given by $\mathbf{k}_f = \mathbf{k}_f^{\text{GLM}|+} - \mathbf{k}_f^{\text{GLM}|+}$, where $\mathbf{k}_f^{\text{GLM}|+} = \mathbf{k}_f^{\text{GLM}}$ for $\mathbf{k}_f^{\text{GLM}} < 0$, and 0 otherwise. This choice of initial conditions consistently gave the best final fits, and for many neurons this was checked for multiple different initial conditions.

We start the optimization procedure by fitting nonlinearities $f_x(\cdot)$ and postsynaptic current terms \mathbf{h}_x , which are fit for both excitation and inhibition terms simultaneously. Then, elements of the excitatory and inhibitory STRFs \mathbf{k}_x are fit by alternating optimization of the excitatory temporal kernel, the inhibitory temporal kernel, the excitatory frequency kernel, and inhibitory frequency kernel. In between each of these optimization steps, internal nonlinearities and postsynaptic current terms are reoptimized. The fitting process is terminated when the LL does not increase any more.

After this optimization, as a final step we measure the spiking nonlinearity using the “histogram method” (Chichilnisky 2001): by measuring the probability of a spike for each value of the sum of the terms inside the spiking nonlinearity (Eq. 1). Because these estimates can be noisy, we describe the spiking nonlinearity with a parametric function, $F(g) = \alpha/\beta \log \{1 + \exp[\beta(g - \theta)]\}$, where α , β , and θ are fit by likelihood optimization. Here, α is the overall slope of the nonlinearity, β determines the sharpness of the transition from zero firing rate, and θ is the threshold. This function empirically resembles the measured spiking nonlinearity quite closely, and its choice has negligible effect on the optimum values for other model parameters.

Dimensionality reduction. Because of the large number of parameters in the two STRFs of the GNM, which contain all combinations of temporal delay and frequency, we approximate each to be frequency-time separable: $\mathbf{k} = \mathbf{k}_t \mathbf{k}_f$, where \mathbf{k}_t denotes the time kernel and \mathbf{k}_f the frequency kernel.

Because excitatory and inhibitory STRFs are both rank-1 (but there are 2 of them), for a direct comparison with the GLM, we allow the GLM STRF using a rank-2 approximation, i.e., as the sum of 2 space-time separable STRFs, $\mathbf{k} = \mathbf{k}_{t1} \mathbf{k}_{f1} + \mathbf{k}_{t2} \mathbf{k}_{f2}$. In fact, going to a higher rank model for the GLM STRF did not significantly improve model predictions ($n = 46$; $P = 0.14$), consistent with earlier findings (Simon et al. 2007).

Further dimensionality reduction is accomplished by representing the temporal kernels in both models by a set of temporal basis functions, $\mathbf{k}_t(t) = \sin\{\pi j[2t/\tau - (t/\tau)^2]\}$ (Keat et al. 2001). Adjusting τ appropriately (130 ms) results in temporal kernels that could accurately describe the range of measured temporal shapes with only eight basis functions.

Finally, to prevent overfitting of postsynaptic current terms, nonlinearities, and frequency kernels, we implement “smoothness regularization” by assessing a penalty for nonsmooth functions. This is accomplished by maximizing the LL of the model with the penalty subtracted, i.e., $\text{LL} - \lambda \Lambda$ (Paninski 2004). For the frequency kernel and the postsynaptic current term, $\Lambda = \sum_i (\mathbf{k}_i - \mathbf{k}_{i-1})^2$ penalizes the slope of these functions. For the internal nonlinearity, we penalize the change of the slope $\Lambda = \sum_i [(\alpha_{i+1} - \alpha_i) - (\alpha_i - \alpha_{i-1})]^2$, where $f_i(x) = \sum_i \alpha_i \xi_i(x)$. λ is chosen separately for each function to ensure smoothly varying kernels.

The effective time kernel and STRF. Temporal processing in the GNM is determined by the temporal kernel \mathbf{k}_t of the receptive field as well as the postsynaptic current term \mathbf{h} . To demonstrate the combined effect of time kernel and postsynaptic current term, we measure the “effective time kernel,” $\mathbf{k}_{\text{eff}} = \mathbf{k}_t * \mathbf{h}$, or the “effective STRF,” $\mathbf{k}_{\text{eff}} = (\mathbf{k}_t * \mathbf{h}) \mathbf{k}_f$, for the GNM. This provides a good approximation to the temporal dynamics generated by each model and allows for direct comparison of their relative latencies.

Entropy as a measure of frequency-tuning width. Because frequency kernels are often not shaped as Gaussians and, for example, might have two peaks, the standard deviation is not a reliable measure of their width. As a result, we used entropy as a measure of their overall width, given by $H(\mathbf{k}_f) = -\sum_i k'_i \log_2 k'_i$, where k'_i is a normalized version of the frequency kernel where all negative values are first set to zero, and the sum over all frequency entries is set to unity. For a Gaussian frequency kernel, this metric yields a value that varies with the logarithm of its width and provides an intuitive measure of log-width of non-Gaussian kernels. Exponentiating the difference between the entropies of the two frequency kernels (Fig. 3A) gives the proportionality between their widths.

Measuring model performance. As a measurement of the model performance, we compute the LL for the model prediction on data that was not used in determining the model parameters, resulting in the reported cross-validated LL (LL_x). We used 83% of the stimulus as training data and the remaining sequences for cross-validation.

Using the LL instead of more traditional performance measures such as r^2 has the advantage that the LL is more sensitive than r^2 and a single spike train is sufficient to obtain a reliable LL. In contrast, r^2 depends on the variance of the spike rate, which for small numbers of repeats is strongly affected by the variability of the number of spike counts in each time bin.

Measuring responses to sound onset and sound offset. We observed different cell types distinguished by their preference for sound onset, sound offset, or continuous sounds. Sound onsets and offsets are most relevant for stimuli such as natural speech, which are composed of distinct spectral elements separated by gaps with little or no sound. To determine the preference of a given neuron to these sound elements, we parse the stimulus into events separated by these gaps and measure event-triggered averages of cell responses. To find onset events, we average the stimulus spectrogram over all frequencies and find all stimulus sequences that have power below a lower threshold, S_1 , for first 50 ms and then, after a 15-ms transition, power above an upper threshold, S_2 , over the following 50 ms. The “onset-triggered average” is given by the average firing rate aligned to these defined onsets (e.g., Fig. 5A). To analyze responses to sound offset, we similarly identify stimulus sequences where the frequency-averaged stimulus is above S_2 for 50 ms and then, after a 15-ms transition, falls below S_1 for the following 50 ms. Averaging across the firing rate of the neuron aligned to this average gives the “offset-triggered average” (e.g., Fig. 5B).

For the purposes of characterizing the distribution of response dynamics across the population of neurons in this study, we used criteria to distinguish neurons that are particularly selective to sound onset, sustained sounds, and sound offsets by comparing their peak response within the 50 ms following sound onset or offset with their “silence rate” and “sound rates.” The silence rate and sound rate are defined by the firing rate during intervals with at least 50 consecutive ms with intensities below S_1 and above S_2 , respectively. Onset cells are defined as onset rate / max (silent rate, sound rate) > 1.7, sustained cells are given by (sound rate > silence rate) and onset rate / max (silence rate, sound rate) < 1.7, and offset cells were defined by offset rate / max (sound rate, silence rate) > 1.2. Neurons that did not meet any of these criteria are considered intermediate cases.

Measuring intensity tuning from natural stimuli. Intensity tuning is typically characterized using pure tones across a range of intensities. However, because in this study we are mainly interested in how intensity tuning is relevant to the processing of natural stimuli, we developed a method to measure intensity tuning directly from the speech data.

This measure of intensity tuning is found by cross-correlating the intensity of the stimulus at the best frequency of the neuron with its response. First, to define a stimulus-intensity for a given neuron at each time point, we smooth the stimulus intensity at the best frequency of the neuron over 20-ms bins. We then associate each stimulus intensity with the observed firing rate delayed by the latency of the neuron (determined by its peak in temporal STRF) and

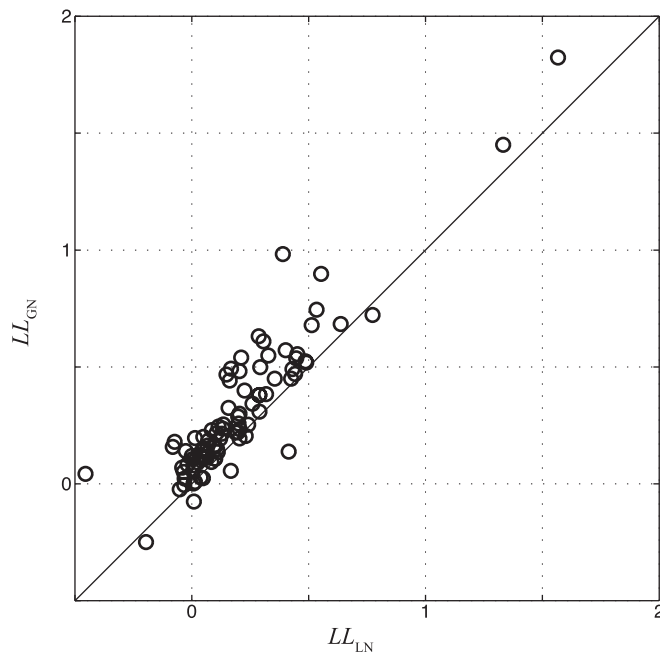


Fig. 2. Comparison of GNM and LN model performance. Each circle represents a neuron in this study ($n = 94$), comparing the log-likelihood (LL) calculated from cross-validation data for the LN model (horizontal axis) and GNM (vertical axis). Nearly all points are above the diagonal line, showing the GNM performs significantly better than the LN model ($P = 4 \times 10^{-12}$) for predicting A1 responses.

smoothed over 20-ms bins. To report a single average firing rate as a function of intensity, we then average the associated firing rate over all instances of that intensity.

RESULTS

To study the interplay of excitation and inhibition in A1 neurons during the processing of complex natural stimuli, we recorded the extracellular responses of 94 single units in passively listening, awake ferrets (David et al. 2009) to continuous human speech from the TIMIT database (Garofolo 1988). We first characterized the representation of sound features in these neurons by their STRF, shown for an example A1 neuron (Fig. 1*B*, top). Because the STRF is a linear measure, its features have an easy interpretation: excitatory (red) regions denote the frequency and relative timing of stimulus features that, on average, result in a high spike rate, and suppressive (blue) regions label the stimulus features associated with lower spike rates. The STRF has proven useful for characterizing basic tuning properties such as frequency tuning and bandwidth.

Overlap between excitation and inhibition in the GNM. However, because it is a linear measure, the STRF cannot detect overlapping excitation and inhibition. This limitation is due to the fact that the STRF only represents the average correlation between stimulus and response, and excitation and inhibition at a given frequency will cancel each other out. As a result, a given spectrotemporal location in the STRF can only show the dominant excitatory or inhibitory component present (i.e., red and blue regions in Fig. 1*B*).

To address this limitation of linear STRF-based models, we use a new approach, the GNM, which explicitly models excitation and inhibition with separate linear STRFs that are

combined nonlinearly to produce the neural response (Butts et al. 2011). The GNM framework models excitation and inhibition as separate nonlinear elements, each composed of a linear STRF that specifies the stimulus tuning of the element (Fig. 1*B*, bottom), and an internal nonlinearity that describes how processing by its STRF nonlinearly influences the response of the neuron (Fig. 1*C*). The two nonlinear elements are linearly combined in a final linear stage before a final spiking nonlinearity is applied to map the combination of excitation and inhibition into a firing rate (Fig. 1*A*, right). This “NLN(N) cascade” is specifically designed to capture the effects of excitation and inhibition on the response of a neuron: given the internal nonlinearities are rectifying, each nonlinear element can only either contribute to an increase in firing rate or decrease in firing rate but not both. Because presumably A1 neurons receive many inputs (excitatory and inhibitory), this model might be viewed as a next-order advance to the single linear STRF-based model, which now considers two pools of neurons instead of one. Because the excitatory and inhibitory STRFs are separately processed by their internal nonlinearity before getting summed together, their effects generally will not cancel at a given location, meaning that overlapping excitation and inhibition can have nontrivial effects on the response of the neuron to complex stimuli and, as we show, can contribute to nonlinear effects such as intensity tuning and nonlinear selectivity to stimulus dynamics.

The GNM reveals largely overlapping spectrotemporal tuning of excitation and inhibition (Fig. 1*B*), as expected from intracellular studies (Tan and Wehr 2009; Wehr and Zador 2003; Wu et al. 2006). Furthermore, it captures exactly how each term is nonlinear, finding that the “internal nonlinearities” for both excitation and inhibition are typically rectifying and saturating (Fig. 1*C*). Although they often share such qualitative features, the differences between these internal nonlinearities can impart important dynamic properties of the responses of the neuron, as described below.

The GNM also describes the observed spike train more accurately than the standard LN model based on the STRF (Fig. 1*D*), with the percent of explained variance (or r^2) of the GNM nearly double that of the STRF-based LN model for

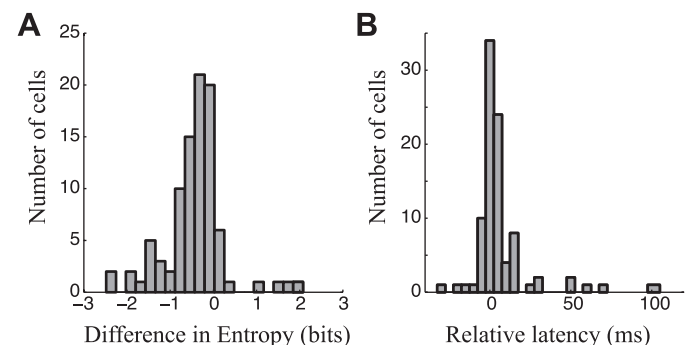


Fig. 3. Comparisons between excitatory and inhibitory STRFs estimated by the GNM. *A*: histogram showing the difference between the widths of excitatory and inhibitory frequency kernels for all neurons in this study ($n = 94$), measured using entropy as a nonparametric measure of broadness of tuning. Nearly all values are negative, demonstrating that the inhibitory frequency kernel of the model is nearly always more broadly tuned than that of excitation. *B*: histogram showing the measured delay between the peak of the effective excitatory and inhibitory time kernels for the same neurons ($n = 94$), demonstrating that inhibition is usually delayed relative to excitation.

the neuron shown [r^2 (GN) = 0.12 and r^2 (LN) = 0.07]. Because of the typical variability of the spike train recorded in awake A1 as well as the number of repeated presentations of the stimulus this allowed (resulting in the relatively low r^2 for

both models), we move to a more sensitive measure of performance, the LL_x (Eq. 1; see MATERIALS AND METHODS), which shows a similar trend as r^2 : LL_x (GN) = 0.40 bits per spike and LL_x (LN) = 0.30 bits per spike. The LL_x is generally a more

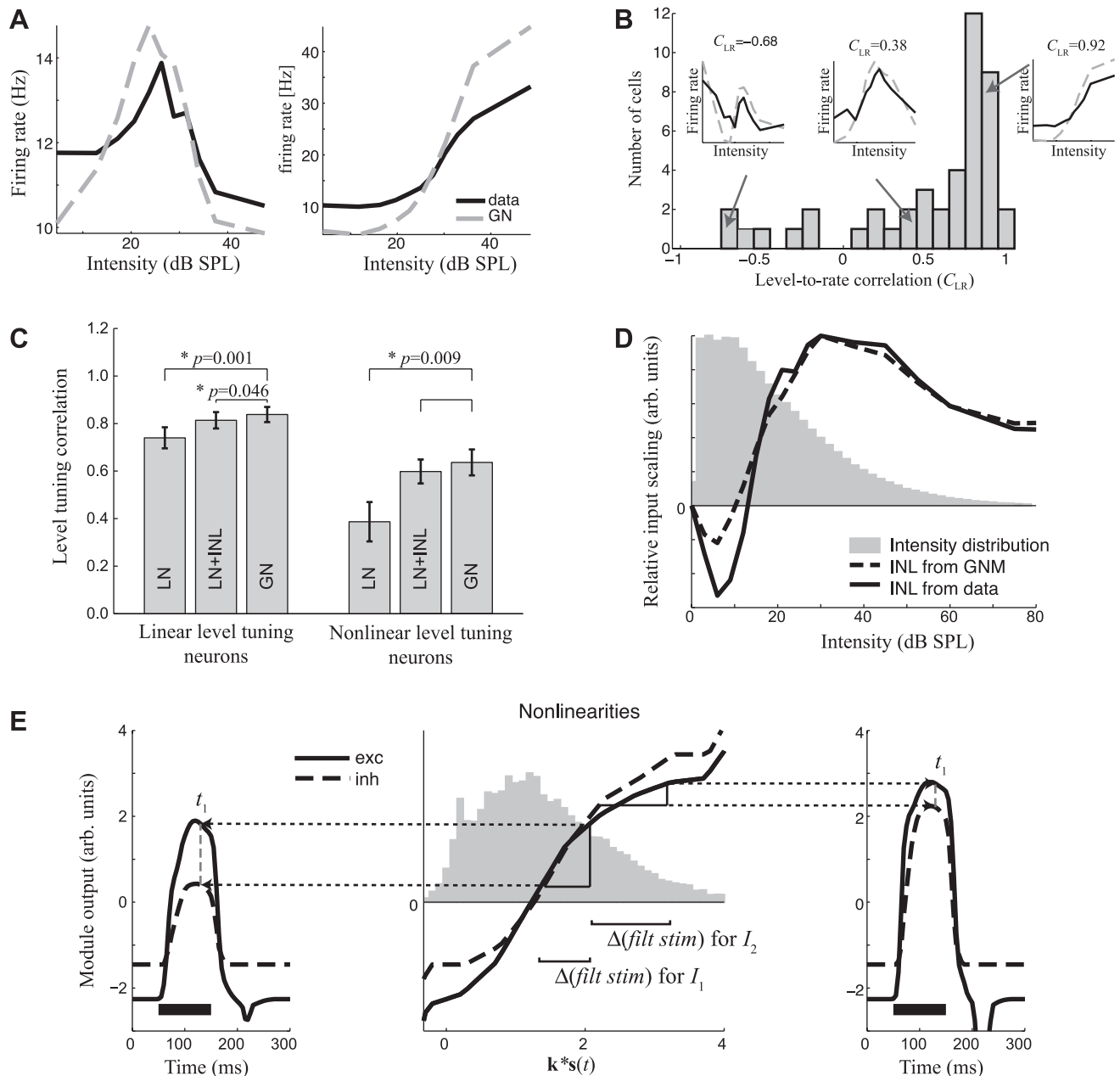


Fig. 4. Intensity tuning through the interplay of excitation and inhibition as predicted by the GNM. **A**: level-tuning curves of 2 example neurons measured in the context of speech stimuli (black). The shapes of these curves are reproduced by level-tuning curves measured from simulated spike trains generated by the GNM (dashed) for each neuron. SPL, sound pressure level. **B**: histogram of the correlation between sound intensity and firing rate across the population of neurons in this study ($n = 94$). This “level-rate correlation” is negative for neurons that respond best to soft sounds (left), small for neurons with nonmonotonic intensity tuning (middle), and close to unity for neurons with monotonic intensity tuning (right). Insets demonstrate particular examples, compared with what is predicted by the GNM, as in **A**. **C**: the mean level-rate correlation coefficient (C_{LR}) across subsets of neurons between the intensity tuning curve of the measured data and the prediction of the different models for neurons with more linear ($C_{LR} > 0.75$) and more nonlinear ($C_{LR} < 0.75$) intensity tuning. **D**: a direct measurement of the input nonlinearity (INL) into an LN model, shown for an example neuron that has a nonmonotonic intensity tuning (black), is another way to characterize intensity tuning. This measure of intensity tuning is also largely reproduced by the output of the GNM (dashed), shown relative to the overall distribution of intensities of the stimulus (gray histogram). **E**: a demonstration of how intensity tuning is generated by the GNM. The outer plots show the excitatory (solid) and inhibitory (dashed) module output for a 100-ms pure tone (black bar) at the preferred frequency of the cell and at 2 different intensities (I_1 ; left; I_2 ; right). At $t_1 = 80$ ms after tone onset (dashed gray), excitation and inhibition differ substantially for the lower intensity, I_1 (left), but much less so for the higher intensity, I_2 (right). The strength of the module outputs depends on the respective nonlinearities (middle), which are shown in relation to the distribution of the generating signal. Because the nonlinearities here are saturating, for a similar difference in the generating signal (middle, horizontal axis), a larger difference in module output (vertical axis) occurs at lower intensity (left) compared with the higher intensity (right). $\Delta(\text{filt stim})$, Change in filtered stimulus.

sensitive measure for model performance than r^2 (Butts et al. 2011) and is particularly useful at the high time resolutions used here (5 ms), which typically require more than five repeated trials to estimate accurately firing-rate based metrics such as r^2 . Using this measure allows us to report accurately model performance across the population of neurons recorded ($n = 94$), which shows significant improvement in the GNM over the STRF-based model (Fig. 2).

Receptive field properties of excitatory and inhibitory elements. Although the GNM uses extracellular data to infer indirectly the excitatory and inhibitory tuning properties, our results are consistent with previous characterizations of the spectral and temporal tuning of inhibition observed in intracellular studies (Sun et al. 2010; Sutter et al. 1999; Tan and Wehr 2009; Wehr and Zador 2003; Wu et al. 2006). In particular, putative excitation and inhibition are usually cotuned, with inhibition and excitation peaking at the same frequency for 44% of the neurons in this study (41/94) and within $\frac{1}{2}$ octave for an additional 35% (33/94).

We also compared the frequency selectivity of each model component using entropy as a nonparametric measure of tuning width (see MATERIALS AND METHODS). Higher entropy corresponds to broader, more distributed tuning, whereas lower entropy signifies sharper, peaked frequency sensitivity. Because of the non-Gaussian shape of many of the frequency-tuning curves, entropy is more consistent as a measure of tuning-curve width than conventional measures such as standard deviation. For 88% (83/94) of the analyzed cells, the inhibitory frequency kernel is broader than the excitatory one (Fig. 3A), and the median difference in entropy corresponds to a 40% increase in the width (measured in octaves) of inhibition relative to excitation. This observation is broadly consistent with previous intracellular studies comparing the widths of excitation and inhibition using pure tones (Wu et al. 2008).

Finally, we compared the latency tuning in the GNM and found a delay between peak excitation and peak inhibition for the majority of cells (Fig. 3B), as also has been previously observed intracellularly (Oswald et al. 2006; Tan and Wehr 2009; Wehr and Zador 2003). However, as described in more detail below, the difference in latency of the STRF itself is just part of the picture, and the overall delay of output of excitation and inhibition in complex stimulus contexts can also be influenced by other factors than the difference in their temporal peaks, including their shape

and rising phase, as well as the dynamics of the stimulus and details of the nonlinear processing itself.

Intensity tuning through the interplay of excitation and inhibition. Although linear models implicitly predict stronger responses for a louder stimulus, a fraction of A1 neurons fire maximally for stimuli at a particular intensity, above which their response decreases with increasing sound level (Sutter and Loftus 2003). Even neurons that are not intensity-tuned will still not respond linearly with intensity and, for example, might saturate or have a different slope at different intensity levels. Such nonlinear intensity tuning of A1 neurons has been a focus of much past work (Sutter and Loftus 2003; Tan et al. 2007; Watkins and Barbour 2011; Wu et al. 2006) and has been explicitly addressed in previous models of A1 neurons (Ahrens et al. 2008b).

Although normally studied in the context of pure tones, this study provided the opportunity to analyze whether intensity tuning is present and relevant in the context of naturalistic stimuli. Intensity tuning can be measured directly from responses of the neuron to naturalistic sounds by averaging the firing rate over brief segments of the speech stimulus with different intensities at the best frequency of the neuron (see MATERIALS AND METHODS). These measurements reveal a diversity of intensity tuning (Fig. 4A), much of which is nonlinear.

We measured the degree of nonlinear intensity tuning using the correlation coefficient between stimulus intensity and resulting firing rate, the level-rate correlation coefficient, C_{LR} (Fig. 4B), with linear relationships having the highest correlation (maximum of 1). Although few cells are fully linear, half have $C_{LR} > 0.75$ and are approximately monotonic and saturating (Fig. 4B, right inset). The other half of the neurons have a diversity of nonlinear intensity tuning, for example with a peak at intermediate intensity (Fig. 4B, middle inset) or with intensity and firing rate not discernibly correlated (Fig. 4B, left inset). The overall prevalence of intensity tuning observed here is consistent with former studies (Clarey et al. 1994; Imig et al. 1990; Phillips and Irvine 1981; Sutter and Schreiner 1995).

Because of the observed intensity tuning, the incorporation of an input nonlinearity that directly scales the intensity of the stimulus (Ahrens et al. 2008a) is able to improve the performance of an STRF-based model ("LN+INL" in Fig. 4C) and better predicts the observed nonlinear intensity tuning. We were surprised to find, however, that the GNM, which has no explicit mechanism for intensity tuning, predicts intensity tun-

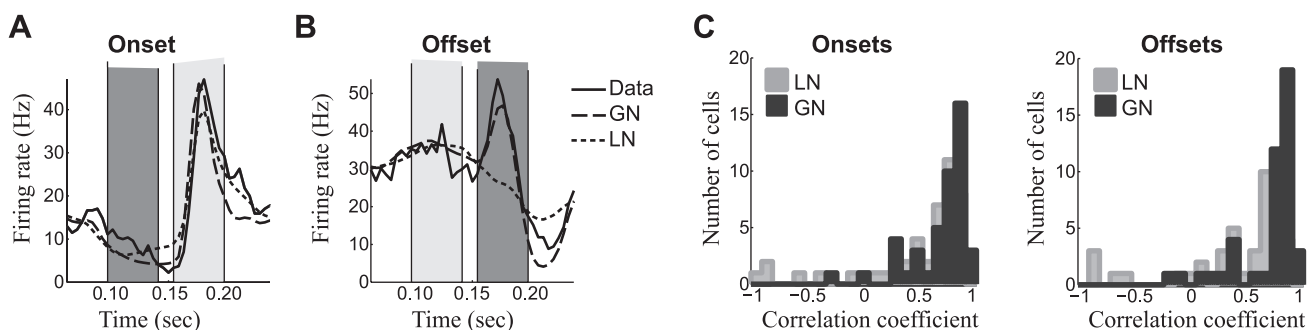


Fig. 5. Responses to stimulus dynamics. *A*: the average response to stimulus onsets of an example neuron that has a strong onset response (solid), compared with that predicted by the GNM (dashed) and LN model (dotted). PSTH and model predictions have been averaged over all stimulus sequences that had low sound intensity during the time interval marked in dark gray followed by high sound intensity during the time interval marked in light gray (correlation coefficients with the data for each model: $C_{LN} = 0.81$, $C_{GN} = 0.87$). *B*: the average response to offsets of an example neuron that has a strong offset response, with similar model comparisons ($C_{LN} = -0.60$, $C_{GN} = 0.90$). *C*: histograms show the correlation coefficient between the event-triggered averages for the measured firing rate and the firing rate predicted by the LN model (gray) and the GNM (black) for onsets (*C*) and offsets (*right*) across the subset of neurons where enough data were available to estimate reliably the event-triggered averages ($n = 45$).

ing just as well as where intensity tuning is explicitly fit. This is demonstrated by comparing the correlation between the level tuning curves calculated from each model with the observed level tuning curves (Fig. 4C). Across the population, the match is significantly better for the GNM than for the LN model (t -test: $P = 0.02$), and this difference is more pronounced for the intensity-tuned neurons ($C_{LR} < 0.75$; Fig. 4C). The emergent property of intensity tuning in the GNM can be further confirmed by generating simulated data using the GNM that is fit to a given neuron. When the LN model with an input nonlinearity is applied to this simulated data, it identifies an input nonlinearity that is very close to that measured from the data (Fig. 4D).

How does the GNM create intensity tuning? Intensity tuning occurs because the relative strength of excitation compared with inhibition differs at different intensities. This balance between excitation and inhibition depends on the internal nonlinearities of each element, combined with the spectrotemporal dynamics of the stimulus. Consider the simple example of the presentation of pure tones, with two different intensities, I_1 and I_2 (Fig. 4E). At a particular time after the stimulus onset (for example, $t_1 = 80$ ms), the stimulus filtered by the excitatory vs. inhibitory STRF is higher for excitation due to the different timing of STRF processing (with inhibition lagging). Because of the shape of the internal nonlinearities, for the lower intensity, I_1 , a smaller difference in filtered stimulus between excitation and inhibition will lead to a large difference in the excitatory vs. inhibitory output (Fig. 4E, *middle* and *left*), resulting in a high firing rate (data not shown). This situation changes at a higher intensity, I_2 , where the filtered stimuli are even further apart but project onto a part of their respective nonlinearities where the slope is much smaller (Fig. 4E, *middle* and *right*). As a result, the temporal delay of the inhibitory STRF processing does not result in a large difference in module output, and the net response is attenuated.

In this way, the nonlinear processing of excitation and inhibition, combined with temporal dynamics of the stimulus, gives rise to nonmonotonic intensity tuning. Such intensity tuning is emergent in the context of the GNM, resulting from the excitatory and inhibitory terms meant to predict the observed spike times. In addition to rediscovering this potential mechanism for intensity tuning (which has been suggested in intracellular studies; Tan et al. 2007; Wu et al. 2006), our work suggests that intensity tuning will also be influenced by stimulus dynamics as well as the presence of particular stimulus features that match the excitatory and/or inhibitory STRFs and thus is likely to play a more complex role in the processing of natural sounds.

Diversity of responses to stimulus dynamics through the interplay of excitation and inhibition. In addition to providing a possible mechanism for nonmonotonic level tuning, nonlinearities associated with excitation and inhibition may also govern how neurons respond to sound onsets and offsets. Onset and offset dynamics are typically studied using isolated sounds, and, as with intensity tuning, the use of a more complex stimulus requires the derivation of new measures of onset and offset tuning. We capture such tuning in the context of speech stimuli by measuring event-triggered average firing rates (Fig. 5, *A* and *B*), where the firing rate of the neuron is averaged over all instances of a sharp increase or decrease in stimulus intensity at the best frequency of the neuron (see

MATERIALS AND METHODS). This reveals neurons that have a strong response to sound onset (Fig. 5A) as well as those that respond to sound offsets (Fig. 5B).

The type of response to sound onset and offset can be explained through the interplay of excitation and inhibition, as captured by the GNM. The degree to which the model can capture this for each neuron is represented by the correlation coefficient between observed and predicted event-triggered averages, shown for the particular cases in Fig. 5, *A* and *B*, as well as across the population of neurons in the study where enough data were taken to estimate them accurately ($n = 45$). This reveals that the GNM predicts onset and offset responses significantly better than the LN model (Fig. 5C; paired t -test: $P = 0.01$ for onsets, $P = 0.001$ for offsets).

Furthermore, the event-triggered average also reveals a third class of neurons that have sustained responses to sounds. In such cases, the response to a particular stimulus event can be higher during a sustained period of the sound compared with the sound onset. For the neurons where we recorded enough data to perform this event-triggered analysis ($n = 45$), 60% had a strong onset response with little sustained activity, and 24% gave strong sustained responses. Of the "onset neurons," 15% ($n = 4$) also had strong offset responses. Finally, there was a single example with high spontaneous firing rate that decreased during sounds. The remaining neurons were intermediate cases (see MATERIALS AND METHODS for specific classification criteria).

Like nonlinear intensity tuning, the prediction of the responses of the model to sound onset and sound offset are influenced by interactions between the temporal selectivity of excitation and inhibition and the shape of their nonlinearities, which is considered for example neurons of each type in Fig. 6.

Response to sound onset. Consider an example cell that strongly responds to sound onset (Fig. 6A). In this case, the GNM time kernels of both excitation and inhibition are biphasic, with a large positive peak followed by a smaller negative peak (Fig. 6D). Such biphasic temporal kernels are naturally selective for sound onset, but in this case the output of this temporal filtering is still significantly broader than the actual response of the neuron (Fig. 6J). For this neuron, additional sharpening comes through the delay between excitation and inhibition. At sound onset, the excitatory filter output rises earlier than the inhibitory output, producing sharp onset responses that are promptly extinguished by the delayed inhibition.

The excitatory and inhibitory nonlinearities can also produce transient responses even without any difference in temporal dynamics. This is illustrated by the excitatory and inhibitory nonlinearities for the example neuron that responds strongly to onset (Fig. 6G), which resemble each other in shape, but the excitatory nonlinearity rises at lower filter output. Thus, as sound intensity increases immediately following onset, excitation will rise sooner due to the sharper rise in excitatory nonlinearity. In this way, the nonlinearities can amplify the time delay between excitation and inhibition at sound onset, leading to a strong onset response (Fig. 6J).

Sustained activity. Figure 6B shows an example cell that demonstrates sustained activity. In this case, the GNM time kernels are nearly monophasic and hence do not produce notable peaks at sound onset or offset. Although, like in the previous example, the increase and peak of the inhibitory time kernel is shifted to later times, there is also a large difference between the

nonlinearities for large filter output. Hence, the nonlinearity amplifies the excitatory response to sounds that match its best frequency compared with the inhibitory response, resulting in sustained activity that is larger than the activity at sound onset.

Response to sound offset. The last case we consider is a neuron that responds to sound offset (Fig. 6C). In this case, there is no explicit delay between the GNM temporal kernels, and thus the filter outputs show little difference between excitation and inhibition at sound offset. However, because the excitatory nonlinearity is amplified compared with that of inhibition, inhibition decreases earlier than excitation, resulting in a peak in the firing rate at sound offset.

DISCUSSION

Here, we present a phenomenological model of A1 responses to naturalistic stimuli that characterizes the interplay of putative excitatory and inhibitory inputs, motivated by numerous intracellular studies demonstrating their importance in shaping neuronal tuning (Sun et al. 2010; Tan et al. 2007; Tan and Wehr 2009; Wehr and Zador 2003; Wu et al. 2006). By incorporating these basic, physiologically motivated mechanisms, the model is able to explain several nonlinear response properties of A1 neurons, including intensity tuning and selectivity to stimulus dynamics. These additional nonlinear properties are not explicitly fit by the model, which only uses observed spike times during naturalistic stimuli. Instead, these properties emerge from the resulting interplay of excitation and inhibition. These observations suggest a unified framework that relates the nonlinear physiological properties of the neuron directly to several previously disparate nonlinear functional properties important for processing natural sounds.

In incorporating specific physiological details into a phenomenological model, the GNM can inform future intracellular experiments, which are relatively low-yield and limited to short experimental durations. The GNM characterization of the putative excitatory and inhibitory tuning is consistent with past observations, including comparisons of their spectral tuning (Wu et al. 2008) and their relative timing (Oswald et al. 2006; Tan and Wehr 2009; Wehr and Zador 2003; Zhang et al. 2003) and mechanisms of intensity tuning (Tan et al. 2007; Wu et al. 2006). Importantly, the fact that these functional properties occur in the context of naturalistic stimuli demonstrates the relevance of intracellular dynamics to behaviorally relevant conditions. In this way, the resulting GNM descriptions also describe expected dynamics of inhibition and excitation in more complex contexts, setting the foundation for further exploration using targeted intracellular studies in this stimulus regime.

Intensity tuning predicted by the GNM. A prominent aspect of nonlinear processing observed in A1 is intensity tuning

(Greenwood and Maruyama 1965; Phillips et al. 1995; Schreiner et al. 1992). This is normally studied in the context of simple tones, but its effects are less clear in more natural contexts such as speech, where sound intensity can vary on both short and longer time scales, for example over the course of individual words (David et al. 2009) or due to distances from sound sources and other aspects of a changing natural environment, respectively. We describe a method for measuring intensity tuning in this more natural context, which not only demonstrates the relevance of intensity tuning to more spectrotemporally complex stimuli, but also provides a new measure to validate how well different models capture intensity tuning. Such measured intensity tuning qualitatively, but not quantitatively, agrees with that measured with pure tones (data not shown). It would not necessarily be expected to agree in detail due to the effect of stimulus context on A1 responses (Asari and Zador 2009) as well as the dependence of excitatory and inhibitory balance underlying intensity tuning on the dynamics of the stimulus (e.g., Fig. 4E).

Although the structure of the GNM has no explicit mechanism to capture intensity tuning, it was able to predict A1 intensity tuning in this natural context better than the linear model and as well as or better than a model with explicit intensity tuning (Ahrens et al. 2008b). The GNM produced this intensity tuning through the separate nonlinearities associated with excitation and inhibition, which led to a changing balance between them as a function of stimulus intensity and temporal dynamics. Such an explanation for nonlinear intensity tuning has been proposed to explain intracellular observations of the different dependence of excitation and inhibition separately on intensity of tones (Tan et al. 2007; Wu et al. 2006). The modeling results presented here underscore the relevance of these observations for more natural contexts and further suggest that tuning to intensity and stimulus dynamics are interrelated. This might be explicitly tested in future experiments by presenting tones at best frequency that are ramped up in intensity at different speeds and either comparing the spike output with that predicted by the GNM or measuring the underlying dynamics of excitation and inhibition directly through intracellular recordings.

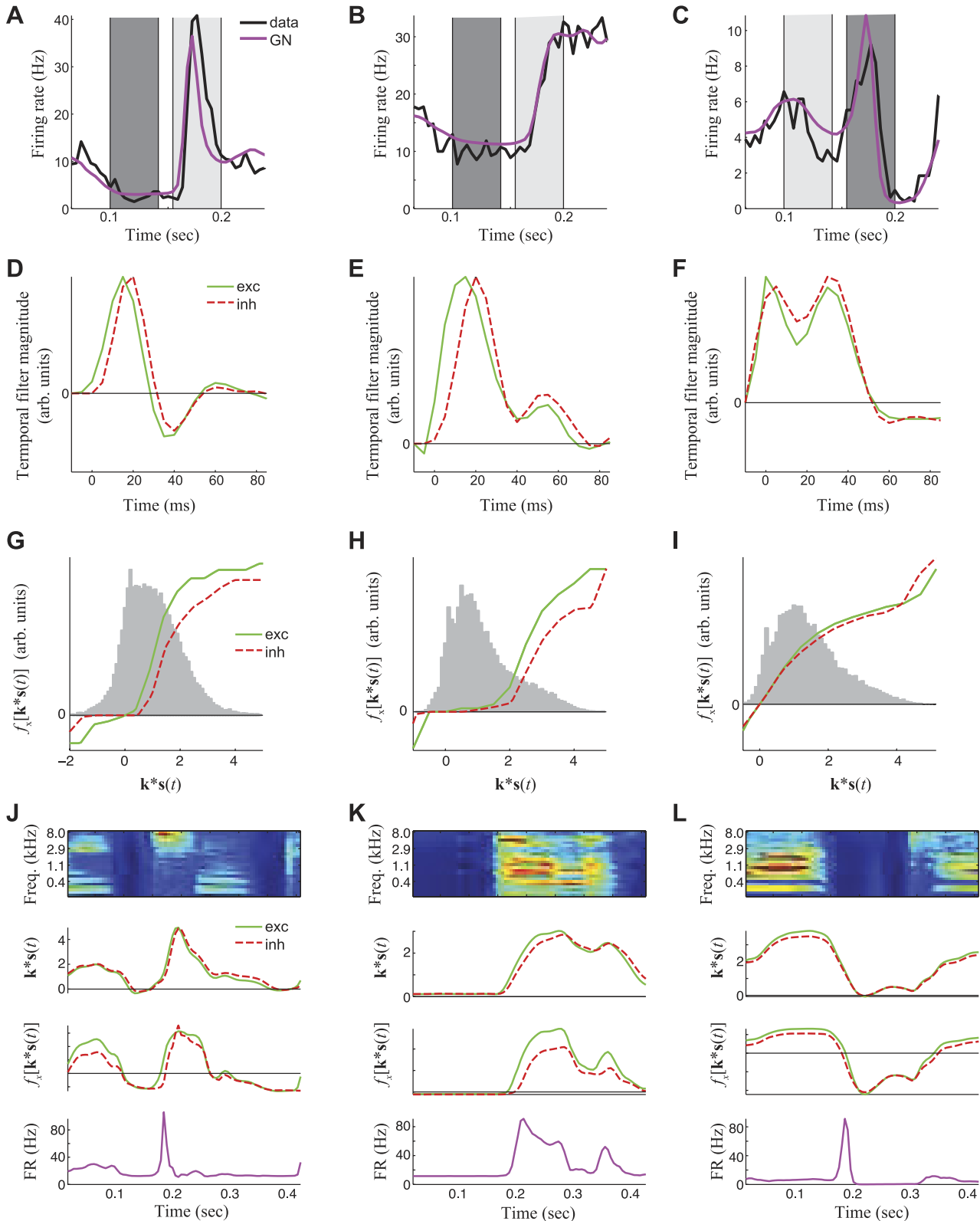
Of course, it is possible that some aspects of intensity tuning are inherited from the inputs to A1 (Tan and Wehr 2009; Wehr and Zador 2003), which can be modeled by an input nonlinearity (Ahrens et al. 2008a). Because such a mechanism, as well other aspects of the multilinear modeling framework suggested by Ahrens et al. (2008a), can be explicitly incorporated into a broader GNM, such possibilities might also be addressed in future modeling studies.

Response dynamics in naturalistic contexts. Speech is different from synthetic stimuli typically used for functional

Fig. 6. The generation of responses to different stimulus dynamics through the interplay of excitation and inhibition as predicted by the GNM. *A*: average PSTH (red) and GNM prediction (blue) at stimulus onset for a neuron with strong onset response. *B*: average PSTH (red) and GNM prediction (blue) at stimulus onset for a neuron with a low onset and strong sustained response. *C*: average PSTH (red) and GNM prediction (blue) at stimulus onset for a neuron with a strong offset response. Time kernels (*D–F*) and nonlinearities (*G–I*) for the GNM that describe the example cells in *A–C*. Here, the gray histograms show the distributions of the filtered stimuli to which the nonlinearities are applied. *J–L*: demonstration of how the temporal kernel and nonlinearities result in the particular response of a neuron to stimulus dynamics, from the sound spectrogram (*top*) to the generating signal, i.e., $k^*s(t)$ (2nd row), filtering by the nonlinearities (3rd row), to the predicted firing rate (*bottom*). *J*: here, an onset causes a peak in the generating signal with a small delay between excitation (green) and inhibition (red). This delay is amplified when processed by the nonlinearity, resulting in a sharp peak of the firing rate (FR) at sound onset. Freq., frequency. *K*: here, the nonlinearity most strongly amplifies excitation over inhibition for high values of the generating signal, resulting in high sustained activity. *L*: at sound offset, excitatory and inhibitory generator functions decrease nearly simultaneously. However, amplification of excitation by the nonlinearity leads to a sharp peak of the firing rate at sound offset.

characterization of A1 [e.g., dynamic random chords (deCharms et al. 1998) and ripples (Kowalski et al. 1996)] in that it is composed of discrete syllables separated by gaps, making the ability of the model to reproduce onset, offset, and sustained responses highly relevant to its overall performance.

The GNM can predict the responses to sound onset and offsets, again through the interplay of excitation and inhibition. Onset neurons can produce transient responses through a relative delay of inhibition compared with excitation, which has been suggested through numerous intracellular studies, both in A1



(Tan and Wehr 2009; Wehr and Zador 2003; Zhang et al. 2003) and throughout sensory cortices (Cardin et al. 2007; Gabernet et al. 2005; Okun and Lampl 2008; Wilent and Contreras 2005). In addition, we found that differences in the nonlinear gain associated with excitation and inhibition can contribute to these responses, even when the latencies of excitation and inhibition are the same.

Only 60% of A1 neurons recorded are clear onset cells, and for neurons exhibiting different dynamics, the GNM revealed a different interplay of excitation and inhibition. In particular, 15% of neurons have a clear offset response, which the GNM described by rebound from inhibition. Such a mechanism has been suggested from intracellular studies (Qin et al. 2007), although more recent work found offset responses arising from a distinct set of inputs (Scholl et al. 2010). The GNM in its current form does not address the possibility of separate offset-tuned inputs but in future work might be modified to incorporate a second set of excitatory inputs that could identify such offset inputs. In either case, as with intensity tuning, the relevance of such mechanisms in predicting the response to natural stimuli further supports the relevance of offset responses.

Comparisons with other nonlinear models. Although sensory neurons are typically modeled through linear methods based on the spike-triggered average, there have recently been new techniques to capture prominent nonlinearities in the auditory system (Ahrens et al. 2008b; Atencio et al. 2009; David et al. 2009), in part due to the recognition that linear models miss significant aspects of A1 processing (Machens et al. 2004).

Two main approaches have been taken to model nonlinear aspects of A1 neuron responses. One approach has been to identify additional stimulus features (in addition to the linear STRF) that contribute to the computation underlying the neuronal response, for example using spike-triggered covariance (STC; Sen et al. 2000) or maximally informative dimension (MID) analysis (Atencio et al. 2008, 2009). Such approaches reliably show that computation in A1 involves higher-order computation on multiple stimulus features, although no constraints are introduced into the type of computation being done with these additional features, often making it less clear what mechanisms this higher-order description is capturing and how they contribute to previously characterized nonlinear processing in A1.

A second approach has taken an opposite tack, by incorporating model components to capture specific nonlinear elements of A1 processing, such as intensity tuning and dependence on stimulus context (Ahrens et al. 2008a). This approach has been cast as a multilinear model, which has the additional advantage that it can be efficiently estimated using linear approaches to optimize model parameters.

The GNM takes a third approach, which, instead of modeling specific nonlinear response properties, introduces nonlinearities to approximate the underlying nonlinear mechanisms known to be present (and likely contribute to the response). It focuses on what is expected to be a dominant nonlinearity in A1 (Dorn et al. 2010; Sun et al. 2010; Wehr and Zador 2003) and is not expected to capture nonlinear processing that occurs over the multiple stages preceding A1. Likewise, by looking for filters coupled to separate nonlinearities that are then added together (in an LNL cascade), it potentially lacks the generality

of discovering relevant stimulus dimensions using overall differences in covariance or maximizing information between spikes and the general distribution as in STC (Schwartz et al. 2006) and MID (Sharpee et al. 2004) analyses. This lack of generality could hurt the GNM performance when considering nonlinear processing of forms not captured by its structure, but it will likely be much more efficient with data when the relevant nonlinearities have the same structure.

Thus, although there is no guarantee that focusing on internal nonlinear mechanisms would necessarily describe the observed response nonlinearities, we found that they did, suggesting that this interplay of excitation and inhibition is important in shaping these nonlinear response properties. Although this is by no means proof that such nonlinear mechanisms are responsible for observed A1 nonlinearities, it does provide a unified framework for linking observable physiology to functional models in complex stimulus contexts. In this sense, by approximating biologically plausible nonlinearities, the GNM provides specific mechanistic predictions for how A1 computation is implemented, which can be directly tested using intracellular recordings.

Conclusion. We present a description of processing in A1 in a natural stimulus context and describe how many previously observed nonlinearities may arise through the interplay of excitatory and inhibitory synaptic inputs. In particular, the interplay of excitation and inhibition explained functional properties that were not explicitly modeled, including nonlinear intensity tuning, and responses to sound onset and offset. As the interaction of excitation and inhibition plays a key role in many brain areas, this model may also lead to more accurate descriptions of neuronal processing outside of A1.

GRANTS

S. V. David and S. A. Shamma were supported by National Institute on Deafness and Other Communication Disorders Grants R01-DC-005779 and K99-DC-010439.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

N.S.-B. analyzed data; N.S.-B. and D.A.B. interpreted results of experiments; N.S.-B. prepared figures; N.S.-B. drafted manuscript; N.S.-B., S.V.D., S.A.S., and D.A.B. approved final version of manuscript; S.V.D., S.A.S., and D.A.B. conception and design of research; S.V.D. performed experiments; S.V.D. and D.A.B. edited and revised manuscript.

REFERENCES

- Aertsen AM, Johannesma PI. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* 42: 133–143, 1981.
- Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci* 28: 1929–1942, 2008a.
- Ahrens MB, Paninski L, Sahani M. Inferring input nonlinearities in neural encoding models. *Network* 19: 35–67, 2008b.
- Asari H, Zador AM. Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J Neurophysiol* 102: 2638–2656, 2009.
- Atencio CA, Sharpee TO, Schreiner CE. Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58: 956–966, 2008.

- Atencio CA, Sharpee TO, Schreiner CE.** Hierarchical computation in the canonical auditory cortical circuit. *Proc Natl Acad Sci USA* 106: 21894–21899, 2009.
- Bizley JK, Nodal FR, Nelken I, King AJ.** Functional organization of ferret auditory cortex. *Cereb Cortex* 15: 1637–1653, 2005.
- Butts DA, Weng C, Jin J, Alonso JM, Paninski L.** Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. *J Neurosci* 31: 11313–11327, 2011.
- Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM.** A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS One* 6: e16104, 2011.
- Cardin JA, Palmer LA, Contreras D.** Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *J Neurosci* 27: 10333–10344, 2007.
- Chichilnisky EJ.** A simple white noise analysis of neuronal light responses. *Network* 12: 199–213, 2001.
- Christianson GB, Sahani M, Linden JF.** The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J Neurosci* 28: 446–455, 2008.
- Clarey JC, Barone P, Imig TJ.** Functional organization of sound direction and sound pressure level in primary auditory cortex of the cat. *J Neurophysiol* 72: 2383–2405, 1994.
- 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.
- David SV, Mesgarani N, Shamma SA.** Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* 18: 191–212, 2007.
- deCharms RC, Blake DT, Merzenich MM.** Optimizing sound features for cortical neurons. *Science* 280: 1439–1443, 1998.
- Dorrn AL, Yuan K, Barker AJ, Schreiner CE, Froemke RC.** Developmental sensory experience balances cortical excitation and inhibition. *Nature* 465: 932–936, 2010.
- Gabernet L, Jadhav SP, Feldman DE, Carandini M, Scanziani M.** Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* 48: 315–327, 2005.
- Garofolo J.** *Getting Started with the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database.* Gaithersburg, MD: National Institute of Standards and Technology (NIST), 1988.
- Greenwood DD, Maruyama N.** Excitatory and inhibitory response areas of auditory neurons in the cochlear nucleus. *J Neurophysiol* 28: 863–892, 1965.
- Imig TJ, Irons WA, Samson FR.** Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. *J Neurophysiol* 63: 1448–1466, 1990.
- Keat J, Reinagel P, Reid RC, Meister M.** Predicting every spike: a model for the responses of visual neurons. *Neuron* 30: 803–817, 2001.
- Kouh M, Sharpee TO.** Estimating linear-nonlinear models using Rényi divergences. *Network* 20: 49–68, 2009.
- Kowalski N, Depireux DA, Shamma SA.** Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 1996.
- Machens CK, Wehr MS, Zador AM.** Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24: 1089–1100, 2004.
- Okun M, Lampl I.** Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat Neurosci* 11: 535–537, 2008.
- Oswald AM, Schiff ML, Reyes AD.** Synaptic mechanisms underlying auditory processing. *Curr Opin Neurobiol* 16: 371–376, 2006.
- Paninski L.** Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15: 243–262, 2004.
- Phillips DP, Irvine DR.** Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to intensity. *J Neurophysiol* 45: 48–58, 1981.
- Phillips DP, Semple MN, Kitzes LM.** Factors shaping the tone level sensitivity of single neurons in posterior field of cat auditory cortex. *J Neurophysiol* 73: 674–686, 1995.
- Qin L, Chimoto S, Sakai M, Wang J, Sato Y.** Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. *J Neurophysiol* 97: 3421–3431, 2007.
- Sahani M, Linden J.** How linear are auditory cortical responses? In: *Advances in Neural Information Processing Systems*, edited by Becker S, Thrun S, and Obermayer K. Cambridge, MA: MIT Press, 2003.
- Scholl B, Gao X, Wehr M.** Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. *Neuron* 65: 412–421, 2010.
- Schreiner CE, Mendelson JR, Sutter ML.** Functional topography of cat primary auditory cortex: representation of tone intensity. *Exp Brain Res* 92: 105–122, 1992.
- Schwartz O, Pillow JW, Rust NC, Simoncelli EP.** Spike-triggered neural characterization. *J Vis* 6: 484–507, 2006.
- Sen K, Wright B, Doupe AJ, Bialek W.** Discovering features of natural stimuli relevant to forebrain auditory neurons (Abstract). *Soc Neurosci Abstr* 551: 2000.
- Sharpee T, Rust NC, Bialek W.** Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16: 223–250, 2004.
- Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA.** Temporal symmetry in primary auditory cortex: implications for cortical connectivity. *Neural Comput* 19: 583–638, 2007.
- Snyder D, Miller M.** *Random Point Processes in Time and Space.* New York: Springer-Verlag, 1991.
- Sun YJ, Wu GK, Liu BH, Li P, Zhou M, Xiao Z, Tao HW, Zhang LI.** Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development. *Nature* 465: 927–931, 2010.
- Sutter ML, Loftus WC.** Excitatory and inhibitory intensity tuning in auditory cortex: evidence for multiple inhibitory mechanisms. *J Neurophysiol* 90: 2629–2647, 2003.
- Sutter ML, Schreiner CE.** Topography of intensity tuning in cat primary auditory cortex: single-neuron versus multiple-neuron recordings. *J Neurophysiol* 73: 190–204, 1995.
- Sutter ML, Schreiner CE, McLean M, O'Connor KN, Loftus WC.** Organization of inhibitory frequency receptive fields in cat primary auditory cortex. *J Neurophysiol* 82: 2358–2371, 1999.
- Tan AY, Atencio CA, Polley DB, Merzenich MM, Schreiner CE.** Unbalanced synaptic inhibition can create intensity-tuned auditory cortex neurons. *Neuroscience* 146: 449–462, 2007.
- Tan AY, Wehr M.** Balanced tone-evoked synaptic excitation and inhibition in mouse auditory cortex. *Neuroscience* 163: 1302–1315, 2009.
- 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.
- Watkins PV, Barbour DL.** Level-tuned neurons in primary auditory cortex adapt differently to loud versus soft sounds. *Cereb Cortex* 21: 178–190, 2011.
- Wehr M, Zador AM.** Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426: 442–446, 2003.
- Wilent WB, Contreras D.** Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat Neurosci* 8: 1364–1370, 2005.
- Wu GK, Arbuckle R, Liu BH, Tao HW, Zhang LI.** Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron* 58: 132–143, 2008.
- Wu GK, Li P, Tao HW, Zhang LI.** Nonmonotonic synaptic excitation and imbalanced inhibition underlying cortical intensity tuning. *Neuron* 52: 705–715, 2006.
- Zhang LI, Tan AY, Schreiner CE, Merzenich MM.** Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* 424: 201–205, 2003.
- Zhang T, Yu B.** Boosting with early stopping: convergence and consistency. *Ann Stat* 33: 1538–1579, 2005.