

Integration over Multiple Timescales in Primary Auditory Cortex

Stephen V. David¹ and Shihab A. Shamma²

¹Oregon Hearing Research Center, Oregon Health and Science University, Portland, Oregon 97239, and ²Electrical and Computer Engineering Department, Institute for Systems Research, University of Maryland, College Park, Maryland 20742

Speech and other natural vocalizations are characterized by large modulations in their sound envelope. The timing of these modulations contains critical information for discrimination of important features, such as phonemes. We studied how depression of synaptic inputs, a mechanism frequently reported in cortex, can contribute to the encoding of envelope dynamics. Using a nonlinear stimulus-response model that accounted for synaptic depression, we predicted responses of neurons in ferret primary auditory cortex (A1) to stimuli with natural temporal modulations. The depression model consistently performed better than linear and second-order models previously used to characterize A1 neurons, and it produced more biologically plausible fits. To test how synaptic depression can contribute to temporal stimulus integration, we used nonparametric *maximum a posteriori* decoding to compare the ability of neurons showing and not showing depression to reconstruct the stimulus envelope. Neurons showing evidence for depression reconstructed stimuli over a longer range of latencies. These findings suggest that variation in depression across the cortical population supports a rich code for representing the temporal dynamics of natural sounds.

Introduction

Substantial information about natural sounds is contained in their envelope (i.e., changes in sound level over time). Correctly perceiving speech requires accumulating information about envelope dynamics over periods ranging from tens of milliseconds, in the case of phonemes (Mesgarani et al., 2008), to many seconds, in the case of sentences (Shannon et al., 1995). Cochlear implant patients rely on these envelope cues for speech perception (Won et al., 2011), and deficits in temporal processing have been linked to impaired language comprehension (Wright et al., 2000). Thus, integration of stimulus dynamics over time represents an essential component of auditory perception (Penner and Shiffrin, 1980).

Studies of primary auditory cortex (A1) have identified an influence of sound history on neural responses lasting many hundreds of milliseconds that could reflect neural strategies for temporal integration (Brosch and Schreiner, 1997; Ulanovsky et al., 2004; Asari and Zador, 2009). However, the mechanisms underlying representations on these timescales, particularly for continually varying natural sounds, remain largely unknown. Accordingly, current computational models of neural processing cannot account for these phenomena. Response latency and du-

ration measured using spectrotemporal models are generally <100 ms in A1 (Kowalski et al., 1996). The limited ability of these models to predict neural responses to natural vocalizations and other stimuli may reflect, in part, their failure to explain temporal integration over longer periods (Machens et al., 2004; David et al., 2009).

In this study, we tested whether synaptic depression, a decrease in presynaptic efficacy from prolonged neurotransmitter release, can account for longer integration times in A1. Synaptic depression is widespread across cortical synapses (Tsodyks and Markram, 1997). Theoretical work has suggested that it may play a role in spatiotemporal integration in the visual system (Chance et al., 1998; Carandini et al., 2002) and a more general computational role in neural circuits (Buonomano and Maass, 2009). We hypothesized that the degree of depression in a neuron's inputs provides information about stimulus history that can be read out by observing the effect of depression on the response to the current stimulus.

We recorded the activity of single neurons in A1 of awake ferrets while presenting a noise stimulus that contained the envelope dynamics of natural vocalizations. We then compared the ability of a model that incorporated nonlinear synaptic depression to explain neural responses to the vocalization-modulated noise against linear (Theunissen et al., 2001; David et al., 2009) and second-order Volterra models (Brenner et al., 2000; Pienkowski et al., 2009) previously used to study representation in A1. The synaptic depression model revealed a biologically plausible diversity of depression effects across neurons and consistently predicted spiking activity as well or better than the other models. Additionally, we found that responses of neurons showing evidence for synaptic depression encoded information about the stimulus envelope over a longer period into the past than the

Received May 27, 2013; revised Oct. 2, 2013; accepted Oct. 24, 2013.

Author contributions: S.V.D. and S.A.S. designed research; S.V.D. performed research; S.V.D. analyzed data; S.V.D. and S.A.S. wrote the paper.

This work was supported by the National Institutes of Health Grant R00DC010439 to S.V.D. and Grant R01DC005779 to S.A.S. and the European Research Council Grant ERC 295603 to S.A.S. We thank D. A. Butts for helpful conversations about experimental design and analysis.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Stephen V. David, Oregon Hearing Research Center, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, MC L335A, Portland, OR 97239. E-mail: davidsv@ohsu.edu.
DOI:10.1523/JNEUROSCI.2270-13.2013

Copyright © 2013 the authors 0270-6474/13/3319154-13\$15.00/0

responses of nondepression neurons. These results are consistent with the hypothesis that feedforward synaptic depression permits auditory cortex to integrate information about natural stimulus dynamics over hundreds of milliseconds.

Materials and Methods

Spiking activity was recorded from a total of 352 single units in primary auditory cortex (A1) of five awake, passively listening female ferrets. All experimental procedures were approved by the University of Maryland Animal Care and Use Committee and conformed to standards specified by the National Institutes of Health.

Experimental procedure

Surgical preparation. Animals were implanted with a steel head post to allow for stable recording. While under anesthesia (ketamine followed by isoflurane) and under sterile conditions, the skin and muscles on the top of the head were retracted from the central 4 cm diameter of skull. Several stainless steel bone screws were attached to the skull, a custom steel post was glued on the midline, and the site was covered with bone cement. After surgery, the skin around the implant was allowed to heal. Analgesics and antibiotics were administered under veterinary supervision until recovery.

Neurophysiological recordings. After animals fully recovered from surgery and were habituated to a head-fixed posture, a small craniotomy (1–2 mm diameter) was opened over A1. Single-unit activity was recorded using tungsten microelectrodes (1–5 M Ω , FHC). One to four electrodes positioned by independent microdrives were inserted into the cortex, and electrophysiological activity was recorded using a commercial data acquisition system (Alpha-Omega). Raw signals were digitized and bandpass filtered between 300 and 6000 Hz.

Spiking events were extracted from the continuous electrophysiological signal using principal components analysis and *k*-means clustering (David et al., 2007). Single-unit isolation was quantified from cluster variance and overlap as the fraction of spikes that were likely to be from a single cell rather than from another cell. Only neurons with $\geq 80\%$ isolation were used for analysis. Increasing the threshold for isolation to 95% reduced the number of units but produced all the same trends observed in the data reported here. Neurons were verified as being in A1 according to by their tonotopic organization, latency, and frequency tuning (Kowalski et al., 1996).

Stimulus presentation was controlled by custom software written in MATLAB (MathWorks). Digital acoustic signals were transformed to analog (National Instruments), equalized to achieve flat gain (Rane), amplified (Rane), and attenuated (Hewlett Packard) to the desired sound level. Stimuli were presented monaurally through an earphone (Etymotic Research) contralateral to the neurophysiological recording site. Before each experiment, the equalizer was calibrated according to the acoustic properties of the earphone insertion.

Acoustic stimuli

Temporally orthogonal ripple combinations (TORCs). Neural best frequency and tuning bandwidth were initially characterized with rippled noise stimuli that allowed rapid, efficient estimation of linear spectrotemporal response functions (STRFs) (Klein et al., 2000). Each TORC set was composed of 30 3-s rippled noise combinations, each of which sampled a narrow range of spectral modulations, spanning a range of 5 octaves (125–4000 Hz, 250–800 Hz or 500–16,000 Hz) with spectral resolution 0–1.4 cycles/octave and temporal resolution 4–48 cycles/s. Typically, two to five repetitions of the TORCs were required for stable STRF estimates.

Vocalization-modulated noise. One experimental limitation to exploring the neural representation of natural sounds is that a complete model of auditory tuning involves many parameters, requiring prohibitively large neurophysiological datasets for fitting (David and Gallant, 2005). To focus on nonlinear temporal integration, we developed a new stimulus with simple spectral structure (bandpass noise), modulated by a temporal envelope from natural vocalizations (see Fig. 1, example). This stimulus permitted fitting models with only a single spectral dimension, maximizing statistical power available to characterize nonlinear tempo-

ral processing while activating the system under naturalistic conditions (Garcia-Lazaro et al., 2006).

Vocalization-modulated noise was generated by first extracting the envelope from a recording of continuous human speech or ferret vocalizations. Envelopes from both classes of vocalization had similar $1/f$ statistics, and we did not observe systematic differences between neurons characterized with either stimulus. The envelope was then applied to bandpass noise that matched the excitatory bandwidth of the neuron within one-half octave, determined from the STRF estimated using TORCs (David et al., 2009). Variants of this stimulus, covering the entire bandwidth of the STRF or even broadband across many octaves, are possible, but in this study we elected to focus on the temporal dynamics of the excitatory frequency band. During each experiment, neural activity was recorded using 4–6 repetitions of 30–40 3-s stimuli. For about half of the auditory-responsive cells ($n = 179$), a second dataset was collected, using 20 repetitions of two additional 3-s stimuli to allow for more exact measurement of prediction accuracy.

Temporal response function analysis

General predictive models are a powerful tool for studying how the brain represents and extracts information from complex natural stimuli (Wu et al., 2006; David et al., 2009). However, the effectiveness of this approach, especially for complex nonlinear models, has been limited by problems of dimensionality. Very general models that make few assumptions about mechanism tend to require a large number of parameters. To evaluate more complex nonlinear models of temporal processing, we simplified the problem by removing spectral variations from the stimulus and thus reducing its dimensionality. This approach allowed us to probe a wide range of nonlinear temporal models that can subsequently be tested in a more general stimulus context.

Linear temporal response function. The linear model used in this study is a simplification of the STRF commonly used to study auditory brain systems (Eggermont, 1993; Kowalski et al., 1996; deCharms et al., 1998; Klein et al., 2000; Theunissen et al., 2001; David et al., 2009). Because the vocalization-modulated noise stimulus was modulated uniformly across frequency bands, neural responses could be described by their temporal response function (TRF), simply a function of time. For a stimulus with a time-varying envelope, $s(t)$, $t = 1 \dots T$, and neuron with instantaneous firing rate, $r(t)$, the linear TRF is defined as the filter $h(u)$, such that,

$$r(t) = h_0 + \sum_{u=0}^U h(u)s(t-u) \quad (1)$$

Each coefficient of h indicates the weight applied to the stimulus envelope at time lag $u = 1 \dots U$ before summing. Baseline response, h_0 , captures the spontaneous firing rate in the absence of a stimulus. Positive values indicate components of the stimulus correlated with increased firing, and negative values indicate components correlated with decreased firing (see Fig. 2 examples).

Second-order temporal response function. This model was a second-order Volterra series generalization from the linear TRF, which accounts for multiplicative interactions between stimuli at different time lags (Eggermont, 1993; Pienkowski et al., 2009). The model is a generalization of the linear TRF, in that the first term is identical to that of Equation 1, as follows:

$$r(t) = h_0 + \sum_{u_1=0}^U h_1(u_1)s(t-u_1) + \sum_{u_1=0}^U \sum_{u_2=0}^U h_2(u_1, u_2) \times s(t-u_1)s(t-u_2) \quad (2)$$

Thus, $h_1(u_1)$ represents the same weights as the linear TRF, and $h_2(u_1, u_2)$ indicates the weight applied to the product of the stimulus at lags u_1 and u_2 .

Synaptic depression temporal response function. The synaptic depression TRF was explicitly designed to capture the nonlinear dynamics caused by depression of synaptic inputs to the neuron being characterized (Tsodyks et al., 1998). This model assumed that depression at a given synapse could be explained by two parameters, the rate of vesicle deple-

tion per presynaptic action potential, v , and the time constant of vesicle recovery, τ . If presynaptic activity is proportional to the stimulus, $s(t)$, then the level of depression at the synapse is as follows:

$$d(t) = d(t-1) + s(t-1)[1 - d(t-1)]v - \frac{d(t-1)}{\tau} \quad (3)$$

bounded at $d = 0 \dots 1$, and the “depressed” stimulus effectively reaching the postsynaptic neuron is the original stimulus scaled down by the current level of depression, as follows:

$$s_d(t) = s(t)[1 - d(t)] \quad (4)$$

To incorporate depression into the TRF framework, we used the nonlinearly transformed stimulus, $s_d(t)$, as the input to a linear filter.

Modeling synaptic depression requires the selecting parameter values (v, τ) in addition to the weights of the linear filter. Rather than estimating the depression parameters explicitly, which like much nonlinear regression is subject to instability, we applied the “kernel trick” to our model (Eichhorn et al., 2004). We modeled the inputs to each TRF as a bank of $i = 1 \dots N$ synapses, each with different depression dynamics (v_i, τ_i). We then projected the stimulus, $s_d(t, i)$, through this bank of filters to produce an overcomplete multichannel expansion of the stimulus, with different nonlinear depression characteristics for each input channel. The depression TRF is then a linear filter applied to the depressed stimulus, as follows:

$$r(t) = h_0 + \sum_{i=1}^N \sum_{u=0}^U h_d(u, i) s_d(t - u, i) \quad (5)$$

In this representation, the only parameters that needed to be estimated were the linear filter parameters, $h_d(u, i)$, associated with each synapse (see Fig. 2 examples). At the same time that it prevents numerical instability, this approach allows neurons to be characterized by multiple synapses with different depression characteristics. The TRF can assign non-zero weights to as many synapses as required to best explain the fit data.

For this study, we used a bank of 12 depressing synapses plus a single nondepressing synapse. Depression strength ranged over three possible values, $v \in \{0.5, 1.5, 2.5\} / \max(s)$ dB SPL⁻¹ (where the units of $s(t)$ are dB SPL) and recovery time constants ranged over four possible values, $\tau \in \{20, 80, 200, 400\}$ ms, making a set of $3 \times 4 = 12$ depression parameter combinations. A simple physiological interpretation of the units of depression strength is difficult, as the original model is based on synaptic vesicle pool size and other parameters that cannot be measured extracellularly (Tsodyks et al., 1998). Normalizing depression strength by the maximum of the stimulus envelope produced consistent performance across neurons and correct units for the model in Equation 3. We tested a larger range of depression strengths (up to 5.0 dB SPL⁻¹) and recovery time constants (up to 2000 ms) and found that TRF fits rarely took advantage of the larger parameter values. Instead, the larger range of parameters produced a slight decrease in TRF predictive power, presumably resulting from model overfitting. This does not rule out the possibility that some A1 neurons might indeed undergo depression with greater strength and time constants, but on average, the functional properties of the population tested in our awake ferret preparation falls in the regimen used here.

The slow recovery dynamics specified by the analytical form of the depression TRF permits a longer memory than the linear and second-order TRFs, which cannot account for stimulus influence at time lags greater than U time bins. To control for the effects of simply increasing temporal integration time in a linear model, we approximated the effects of synaptic depression using a bank of linear, exponentially decaying filters,

$$s_i(t, i) = \sum_{u=0}^{\infty} s(t - u) \exp(-u/\tau_i), \quad (6)$$

with variable time constants $\tau_i \in \{5, 10, 20, 40, 80, 160, 320, 640\}$ ms. The filtered stimulus, $s_i(t, i)$ was then input to a linear filter with the same form

as the depression TRF in Equation 5. For large τ_i , this transformation had the effect of low-pass filtering and making stimulus information hundreds of milliseconds before the current time available to a linear filter without requiring a larger maximum time lag, which can worsen fitting noise (see below).

Data preprocessing. For each neuron, all of the models defined above were estimated using the same dataset. Before TRF estimation, data recorded from a neuron were divided into separate estimation and validation segments. To avoid bias from overfitting (especially because the different models contained different numbers of parameters), all model parameters were estimated using the estimation segment and subsequently tested with the validation segment. For neurons presented the 6 s, 20-repeat stimulus, these data were reserved for validation. For the remaining neurons, 7.5% of the total dataset was reserved for validation. Because the 20-repeat data collected for many neurons represented a subset of the data especially designed for prediction noise ceiling calculation (Sahani and Linden, 2003), we elected not to perform a full cross-validation procedure in which separate TRFs were estimated after excluding different validation datasets. A full cross-validation procedure should produce slightly more accurate models but would be unlikely to change the results dramatically (David et al., 2007).

Stimulus and response data were binned at 5 ms, and maximum time lag, U , was set to 100 ms for all analyses, except controls that explicitly varied these values. The stimulus envelope was measured by full-wave rectifying the stimulus waveform, interpolating between local maxima, smoothing to the temporal bin size of the TRF, and applying log compression. Neural responses were computed as the instantaneous firing rate in each time bin, averaged across trials.

Model estimation by boosting. For all models, TRF parameters were estimated by boosting or coordinate descent (David et al., 2007, 2009). Previously, in the context of auditory STRF analysis, boosting has been used only to estimate linear models. Both the second-order and depression TRF models can be cast as linear models if the appropriate nonlinear transformation is applied to the stimulus before fitting a linear filter (Eichhorn et al., 2004). Thus, after nonlinear projection into the appropriate space (second-order stimulus outer product and synaptic depression filter bank, respectively), we could use the same boosting algorithm to estimate TRFs for all three models. We confirmed that boosting accurately recovers the temporal filter properties of simulated neurons that undergo and do not undergo synaptic depression (see Fig. 2). After fitting all TRFs, a static rectification nonlinearity was fit to the TRF output to account for spike threshold (David et al., 2009).

Model validation. Each TRF was evaluated based on its ability to predict the time-varying peristimulus time histogram (PSTH) response in the reserved validation dataset. Prediction accuracy was measured as the correlation coefficient (Pearson’s r) between the predicted and observed PSTH, binned at 5 ms (David et al., 2007). Because some error resulted from uncertainty in the observed PSTH due to finite stimulus repetitions, we also measured prediction accuracy after correcting for this sampling limitation (Sahani and Linden, 2003). This correction improved prediction scores but had no influence on the relative performance of the different models.

Maximum a posteriori reconstruction

Stimulus reconstruction provides a tool for measuring stimulus information encoded by a population of sensory neurons (Bialek et al., 1991; Mesgarani et al., 2009). Most previous reconstruction methods applied to complex natural sensory data have used linear decoding because of its robustness and tractability. However, to characterize stimulus information that was not explicitly available to a linear decoder, we used a model-free, *maximum a posteriori* method. The neural representation was modeled using a population rate code in that neural firing rates measured over a relatively long time window (100 ms) were used to reconstruct the stimulus envelope with finer temporal resolution (5 ms).

The reconstruction procedure was designed to determine the most likely stimulus, $s(t - \tau)$, at time lag, τ , from the response of N neurons at time t , $\mathbf{r}(t) = [r_1(t) \ r_2(t) \ \dots \ r_N(t)]$ and knowledge of the stimulus prior probability. For the purpose of the probability analysis, the stimulus and response at each time point are binned at discrete levels. According to

Bayes' theorem, the probability of a stimulus, $s(t - \tau)$, given the response, $r(t)$, can then be written as follows:

$$P(s(t - \tau)|r(t)) = \frac{P(r(t)|s(t - \tau))P(s(t - \tau))}{P(r(t))} \quad (7)$$

If we make the simplifying assumption that neural responses are independent, then,

$$P(s(t - \tau)|r(t)) = \prod_{i=1}^N \frac{P(r_i(t)|s(t - \tau))P(s(t - \tau))}{P(r_i(t))} \quad (8)$$

and most likely stimulus is

$$S_{ML}(t - \tau) = \operatorname{argmax} \prod_{i=1}^N P(r_i(t)|s(t - \tau))P(s(t - \tau)) \quad (9)$$

Because the marginal, $P(r)$, is independent of s , the denominator of Equation 8 drops out of the maximum computation. Finding the most likely stimulus then simply requires finding the peak of the one-dimensional distribution in Equation 9.

For stimulus reconstruction, we focused on the subset of neurons that had been presented stimuli with identical auditory envelopes ($n = 258$). To test whether neurons showing evidence for synaptic depression encoded stimuli with longer memory, we selected the 57 neurons from this subset that showed a significant improvement in prediction power for the synaptic depression TRF over the linear TRF ($p < 0.05$, jackknifed t test). We compared the accuracy of reconstructions by this group to those by an equal-sized group with the same average prediction accuracy but no difference between depression and linear TRFs. For reconstructions from simulated responses, we used the TRF estimates for those groups of neurons to predict responses to the stimulus (Eqs. 1 and 5, respectively, for linear and depression TRFs). The identical reconstruction analysis was then performed using the simulated responses.

Data preprocessing. The stimulus envelope was binned at 5 ms and discretized into 20 uniform sound levels. The response of each neuron at each time was computed by averaging the spike rate over the preceding 100 ms time window and discretizing into 7 uniform bins. Then at each moment in time, a vector composed of the neural population spike rate was used to reconstruct the preceding stimulus envelope on a 5 ms time-scale. We experimented with different response windows (10–200 ms) and discretization schemes. Changing these parameters slightly diminished differences in the reconstruction between depression and nondepression subpopulations but overall produced the same effects.

Reconstruction validation. Reconstructions were performed using 20-time jackknifing, in which 95% of the data were used for estimating response probabilities, $P(r_i|s)$, and the decoding was performed on the remaining 5%. Reconstruction accuracy was measured as the normalized mean square error between actual and reconstructed stimulus. Significant reconstruction accuracy was evaluated using a jackknifed t test.

We also fit a linear decoder (Bialek et al., 1991; Mesgarani et al., 2009) using the same data and observed similar trends, but performance was worse at longer time lags for both groups of neurons. Thus, some of the long-latency information is not encoded in a form explicitly available to linear decoders.

Results

A diversity of temporal response dynamics in A1

We recorded the activity of 352 single units in A1 of awake, passively listening ferrets during presentation of vocalization-modulated noise (Fig. 1A). The noise stimulus was constructed by applying the temporal envelope of continuous speech or ferret vocalizations to bandpass noise matched to each neuron's preferred frequency. Because spectral energy changed coherently within the stimulus noiseband, the stimulus probed the cortical representation of stimuli with complex, naturalistic temporal dy-

namics but could also be described as a relatively simple one-dimensional function of time (Fig. 1B).

In A1, neural responses to the vocalization-modulated noise were generally robust and reliable over repeated trials. The dynamics of the peri-stimulus time histogram (PSTH) response, however, varied substantially across neurons (Fig. 1C). Some neurons closely followed the stimulus envelope over the entire 3 s noise segment (Fig. 1C, bottom). Other neurons tended to respond only to the onset of each syllable (Fig. 1C, middle), whereas others responded strongly only to the onset of each 3 s stimulus epoch (Fig. 1C, top). Because the stimulus noiseband was matched to the excitatory tuning of each neuron, the differences between PSTHs reflect variability in the neurons' temporal response properties.

Modeling synaptic depression during temporal processing

To determine the source of the variability in PSTHs among neurons, we estimated the temporal response function (TRF) for each neuron, which describes the relationship between the stimulus envelope and the time-varying neural response. Classically, functional characterization of auditory neurons has used the linear STRF, which finds the best linear weighted sum of stimulus spectral channels and time lags to predict the instantaneous neuronal firing rate (Eggermont, 1993; Kowalski et al., 1996; deCharms et al., 1998; Klein et al., 2000; Theunissen et al., 2001; David et al., 2009). In the case of vocalization-modulated noise, the STRF can be simplified to have a single spectral dimension, a linear TRF that provides a characterization of temporal filtering properties in the band around the neuron's best frequency (Fig. 2A).

The linear TRF, although useful for characterizing basic temporal filtering properties, cannot capture some of the temporal dynamics observed in the neural responses (David et al., 2009). It is well known that neurons adapt to sustained stimulation, and it has been proposed that depression of input synapses contributes to this adaptation (Tsodyks and Markram, 1997; Ulanovsky et al., 2004). We simulated the response of a neuron with a single excitatory, 25 ms latency depressing synapse to vocalization-modulated noise and estimated a linear TRF from the response (Fig. 2B). The resulting fit showed distinctly different features from the underlying simulation. The linear TRF had a bimodal, positive-negative time course typical of linear filters measured in auditory cortex (David et al., 2009), despite the fact that the synaptic weights in the simulation were exclusively positive. In addition, the onset latency of the linear TRF was shorter than the actual latency of the simulated neuron (Fig. 2B, positive gain at 15 ms time lag vs actual 25 ms onset latency).

We compared the linear TRF to a nonlinear depression TRF that accounts for the effects of synaptic depression before the linear filtering stage. The strength of depression and its recovery time can vary substantially across neurons (Tsodyks and Markram, 1997), and the depression model accounts for this diversity by passing the stimulus envelope through a filter bank that mimics synapses with different depression properties (Fig. 2A). The depression TRF is then a weighted sum of the output of all these filters, with weights chosen to best predict the response of the neuron (Fig. 2B). When data from the same simulated neuron were fit using this model, the resulting TRF captured the correct excitatory tuning and latency of responses (Fig. 2C). The depression TRF was able to accurately characterize temporal response properties for depression neurons with recovery time constants up to 400 ms, limited only by the maximum time constant included in the depression model (see Materials and Methods).

Longer time constants could be incorporated into the model but did not benefit analyses on the data used in this study (Fig. 6).

The depression TRF subsumes the linear TRF in that one input channel does not undergo depression and simply repeats the input stimulus. For a neuron that does not undergo depression, parameter values in the depression TRF should be zero, except for parameters associated with the nondepressing synapse. We observed this behavior when we simulated the response of a nondepressing neuron with an excitatory response (25 ms latency) followed by inhibition (35 ms latency; Fig. 2D). Thus, the depression TRF can reveal distinct temporal response dynamics (Figs. 2C vs D) in neurons with linear TRFs that appear very similar (Figs. 2B vs D).

Synaptic depression can explain nonlinear response dynamics in A1

We compared linear and depression TRFs estimated from the A1 single unit data. Here, we consider three representative examples. Some neurons showed striking differences between the linear and depression TRF. One neuron gave a strong transient response to the onset of the vocalization modulated noise (Fig. 1C, top row). The linear TRF for this neuron indicated a very short latency excitatory response (5–15 ms), followed by a slower inhibitory response (Fig. 3A). This bimodal filter shape was characteristic of linear TRF estimates for simulated neurons that underwent depression (Fig. 2B). When the linear TRF was used to predict the PSTH response in a validation dataset that was not used for estimation, it was able to approximately predict epochs when the neuron responded (compare black and green curves, Fig. 3A, bottom, prediction correlation, $r = 0.46$). However, it could not predict the strong response at stimulus onset, and it tended to overpredict responses later in the stimulus.

The depression TRF for the same neuron showed only a very weak inhibitory weight at its linear, nondepressing synapse (Fig. 3A, top right). Instead, it showed strong excitatory responses at a number of synapses, with recovery time constants as long as 400 ms. The depression TRF was able to much better predict relative response strength throughout the stimulus (Fig. 3A, bottom, red curve), with a significantly greater correlation between the predicted and observed PSTH ($r = 0.64$, $p < 0.05$, jackknifed t test).

Other neurons showed evidence of synaptic depression with more rapid dynamics. These neurons tended to respond to the onset of each high-energy epoch in the stimulus (Fig. 1C, middle rows). In one such example, the linear TRF was strongly bimodal, with a very short latency excitatory response (Fig. 3B). The de-

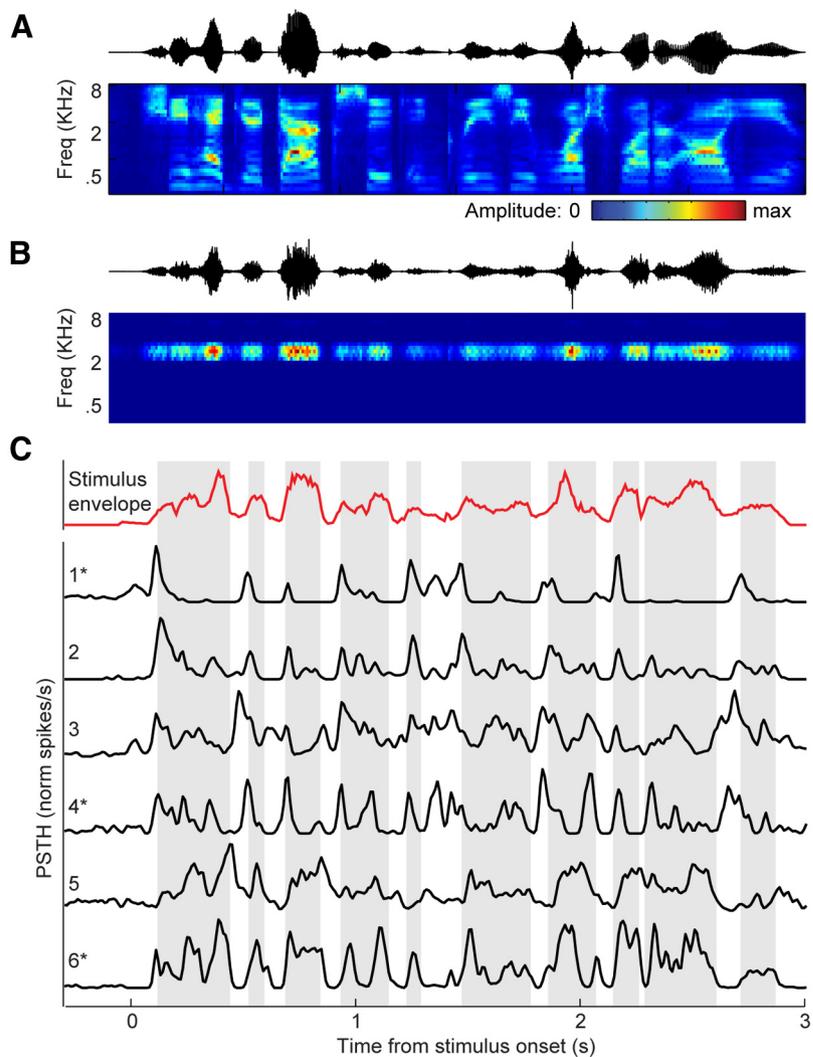


Figure 1. *A*, Example waveform (top) and corresponding spectrogram (bottom) of continuous speech. In the spectrogram, sound energy at a particular frequency (vertical axis) over time (horizontal axis) is indicated by color, where red represents the most energy and blue represents the least. As is typical for speech, syllables occurring approximately once every 300–400 ms are identified by coherent increases in stimulus energy across frequencies, separated by silent periods with very little energy. *B*, Waveform and spectrogram of speech-modulated noise, plotted as in *A*. The temporal envelope of this stimulus matches that of the speech in *A*, but the spectral content is uniform bandpass noise between 2 and 3 kHz. *C*, Diversity of peri-stimulus time histogram (PSTH) responses of six example A1 neurons to a stimulus with the temporal dynamics in *B*. The stimulus envelope is plotted in red in the top row. Periods with greater than average power are shaded in gray. For neurons with linear response properties, the PSTH (black lines) should be coherent with the stimulus envelope, with consistently higher firing rates throughout the gray-shaded areas (bottom rows). Some neurons show a persistent decrease in firing after a strong response to the beginning of the stimulus (top rows) or transient responses to the onset of each syllable (middle rows), indicating nonlinear temporal integration. Data from neurons marked with asterisks appear in Figure 3.

pression TRF found non-zero weights for depressing synapses with a relatively short recovery time constant, and it predicted neural responses significantly better than the linear TRF (linear $r = 0.39$, depression $r = 0.54$, $p < 0.05$, randomized paired t test). Although the linear TRF was able to predict slow modulations in the PSTH, the depression TRF was much better at capturing the rapid dynamics of the transient responses.

The responses of other neurons largely followed the stimulus envelope (Fig. 1C, bottom) and showed little influence of synaptic depression. In a final example, the linear TRF showed an excitatory response 10–20 ms after the stimulus (Fig. 3C). The depression TRF had very similar properties to the linear TRF, with large weights for the nondepressing synapse and much smaller weights for depressing synapses. In this case, the predic-

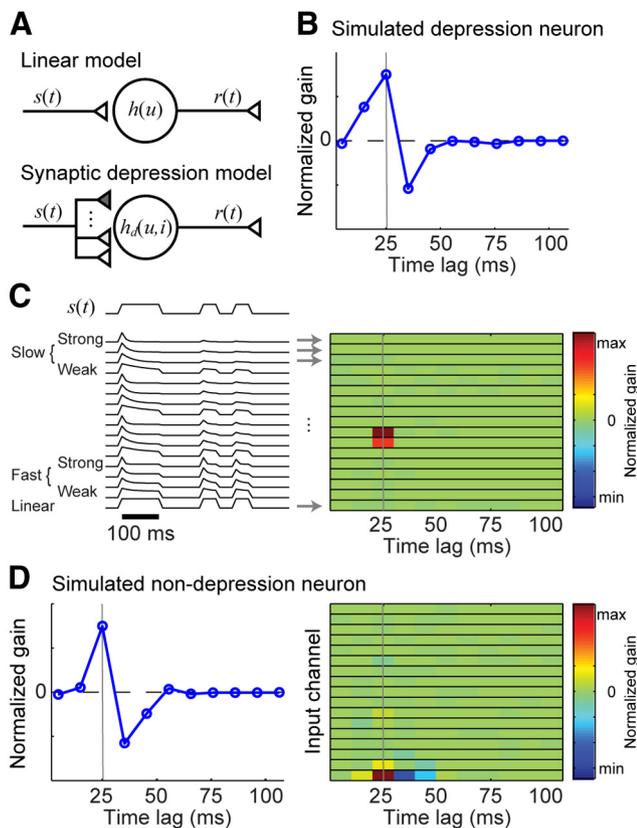


Figure 2. *A*, Schematic of linear and synaptic depression temporal response functions (TRFs) for characterizing vocalization-modulated noise data. For the linear TRF (left), the stimulus can be thought of as a presynaptic spike train with time-varying rate proportional to the stimulus envelope. The TRF, then, mimics the synaptic potential after each input spike with a sequence of weights, $h(u)$, across time lags. The synaptic depression TRF (right) is a generalization of the linear TRF, in which the same input arrives at multiple synapses, each with different depression properties. The depression TRF assigns separate weights to each of these inputs. *B*, Linear TRF estimated for a simulated neuron with a single excitatory input synapse that undergoes depression with a 150 ms recovery time constant. Circles indicate parameter weights estimated from the data. Positive or negative values (vertical axis) indicate a relative increase or decrease, respectively, in output spiking at a given time lag (horizontal axis) after positive stimulus energy. In attempting to account for the dynamics of the depressing synapse, the linear TRF has shorter onset latency (15 ms) than the actual simulation (25 ms, gray line) and a late inhibitory response. *C*, Depression TRF estimated for the same simulated neuron. Left, Each row schematically illustrates the time course of synaptic output for a stimulus consisting of a sequence of square pulses (top row). Strong, long-lasting depression results in very little neurotransmitter release after the onset of the first pulse (upper rows). Shorter-lasting depression results in transient responses after the onset of each pulse (middle rows). A nondepressing, linear synapse simply repeats the input (bottom row). The heatmap at right shows estimated synaptic weights. Red or blue regions represent increases or decreases, respectively, in output spiking after input at the corresponding synapse, with latency on the horizontal axis. The positive weights for the 80 and 200 ms synapses straddle the dynamics of the underlying 150 ms synapse of the simulated neuron. Unlike the linear TRF, the depression TRF captures the correct response latency and absence of late inhibitory inputs. *D*, For a simulated neuron with input (short-latency excitation followed by longer-latency inhibition) that does not undergo depression, the linear TRF (left) accurately captures the neuron's filter properties. The depression TRF (right) finds similar properties, with non-zero weights for the nondepressing synapse and the same time course as the linear TRF (bottom row of heatmap).

tive power of the linear and depression TRFs was the same (linear $r = 0.74$, depression, $r = 0.74$, $p > 0.1$, randomized paired t test). Both models also predicted very similar PSTHs.

We compared performance of the TRF models over the entire set of A1 neurons in our study (Fig. 4A). Across all the neurons studied, either the linear or the synaptic depression TRF was able to predict responses in the validation dataset significantly above

chance in 307 of 352 (87%, jackknifed t test, $p < 0.05$). Among these responsive neurons, the synaptic depression TRF performed significantly better than the linear TRF for 71 of 307 neurons (23%, $p < 0.05$, jackknifed t test), whereas the linear TRF did not perform significantly better for any neuron. Even for neurons without a significant change for the synaptic depression model, there was a clear trend toward improvement. For the entire set of responsive neurons, the average prediction correlation for the synaptic depression TRF (mean $r = 0.33$) was significantly greater than that of the linear TRF (mean $r = 0.29$, $p < 0.0001$, randomized paired t test).

To obtain an absolute measure of model performance, we computed the fraction of response variance explained (r^2), after normalization for the effects of noise from finite stimulus repetition in the validation data (Sahani and Linden, 2003). After normalization, the average fraction of response variance explained was $r^2 = 0.37$ for the linear TRF and $r^2 = 0.46$ for the depression TRF (Fig. 4B; $n = 179$ neurons with adequate repeats of the validation stimulus, $p < 0.0001$, randomized paired t test). Thus, a TRF incorporating synaptic depression shows substantially improved performance over the linear model, with an average 24% increase in variance explained.

The fact that the depression model contained a larger number of parameters than the linear model (260 vs 20) raised the possibility that its improved performance simply reflected the generic ability of a more complex model to describe nonlinear responses, rather than specifically reflecting the dynamics of synaptic depression. To test for this possibility and to compare the performance of the depression TRF to a more state-of-the-art model, we also evaluated a second-order TRF on the same data (Pienkowski et al., 2009). The second-order model used a Volterra series expansion of the stimulus in time to account for multiplicative interactions at different latencies. It required slightly more parameters (400) than the synaptic depression TRF, and thus was approximately matched in complexity. The average second-order TRF did perform significantly better than the linear TRF by $\sim 9\%$ (mean fraction of normalized variance explained, $r^2 = 0.40$, $p < 0.001$; Fig. 4B), but it still performed significantly worse than the synaptic depression TRF ($p < 0.0001$). Thus, a model explicitly accounting for synaptic depression was better able to explain time-varying responses in A1 than a more generic nonlinear model with similar complexity.

Because the synaptic depression TRF incorporates long time constants, one possible explanation for its improved performance is that it integrates stimulus information over longer time lags than either other TRF. To test for simple integration time effects, we performed three control analyses. First, we increased the maximum time lag of the linear TRF to 200 rather than 100 ms. This change did not significantly change performance and actually slightly decreased predictive power, presumably because of the increased number of model parameters (mean normalized $r^2 = 0.36$). Second, we fit all three models with coarser 20 ms rather than 5 ms time bins, thus increasing integration time by a factor of 4 without changing the number of model parameters. In this case, we observed an overall apparent improvement in performance for all models that resulted from having temporally smoothed the data (mean normalized $r^2 = 0.41, 0.45, 0.52$, for linear, second-order and depression TRFs, respectively). However, this manipulation did not change the relative performance of the different models (second-order TRF +11%, depression TRF +26% relative to the linear TRF), nor did it change the time constants of the depression TRF fits. Finally, we substituted a bank of linear, decaying exponential filters into the depression

TRF (see Materials and Methods, Eq. 5) and fit a linear approximation to the depression TRF using the same procedure as for the other models. This change caused a slight but insignificant decrease in performance relative to the linear TRF (mean normalized $r^2 = 0.34$), again presumably reflecting the larger parameter count. Thus, the improved performance of the depression TRF likely reflects its distinct analytical formulation rather than simply its longer memory.

Synaptic depression model reveals biologically plausible temporal tuning

To characterize the dynamics revealed by the linear and depression TRF estimates, we computed the average TRF across neurons. Neurons were grouped according to whether the synaptic depression model significantly improved predictive power (depression neurons, Fig. 4A, filled circles, $n = 71$) or did not (nondepression neurons, $n = 236$). The average linear TRF differed substantially between the two groups. For nondepression neurons, the linear TRFs were largely positive, with onset latencies of 10–20 ms (Fig. 5A, left). In contrast, depression neurons produced bimodal linear TRFs that changed between positive to negative at short to long latencies, respectively (Fig. 5B, left). Particularly striking about this subset of linear TRFs is that their onset latency appears very short (<10 ms), earlier than that considered plausible for A1 neurons (Bizley et al., 2005).

To compare linear TRF properties with those of the depression TRF, we computed the average weights for the synapse with highest overall gain (SD across time lags) that underwent depression (i.e., any row of the depression TRF except the bottom; Fig. 2C) and for the nondepressing synapse (bottom row; Fig. 2C). As expected, the average TRF of the nondepression neurons was similar to that of the average linear TRF, concomitant with the example neuron for which both models performed similarly (Fig. 3C). The depressing synapse showed only a weak response, and the nondepressing synapse showed similar dynamics to the linear TRF (Fig. 5A, right). In contrast, the average TRF for depression neurons showed a strong positive response for the depressing synapse and a negative response for the nondepressing synapse (Fig. 5B, right). These TRFs had latencies closer to what would be expected for A1 neurons (10–15 ms) than the average linear TRF. Slight deviations from zero remain at extremely short and long latencies, even in the

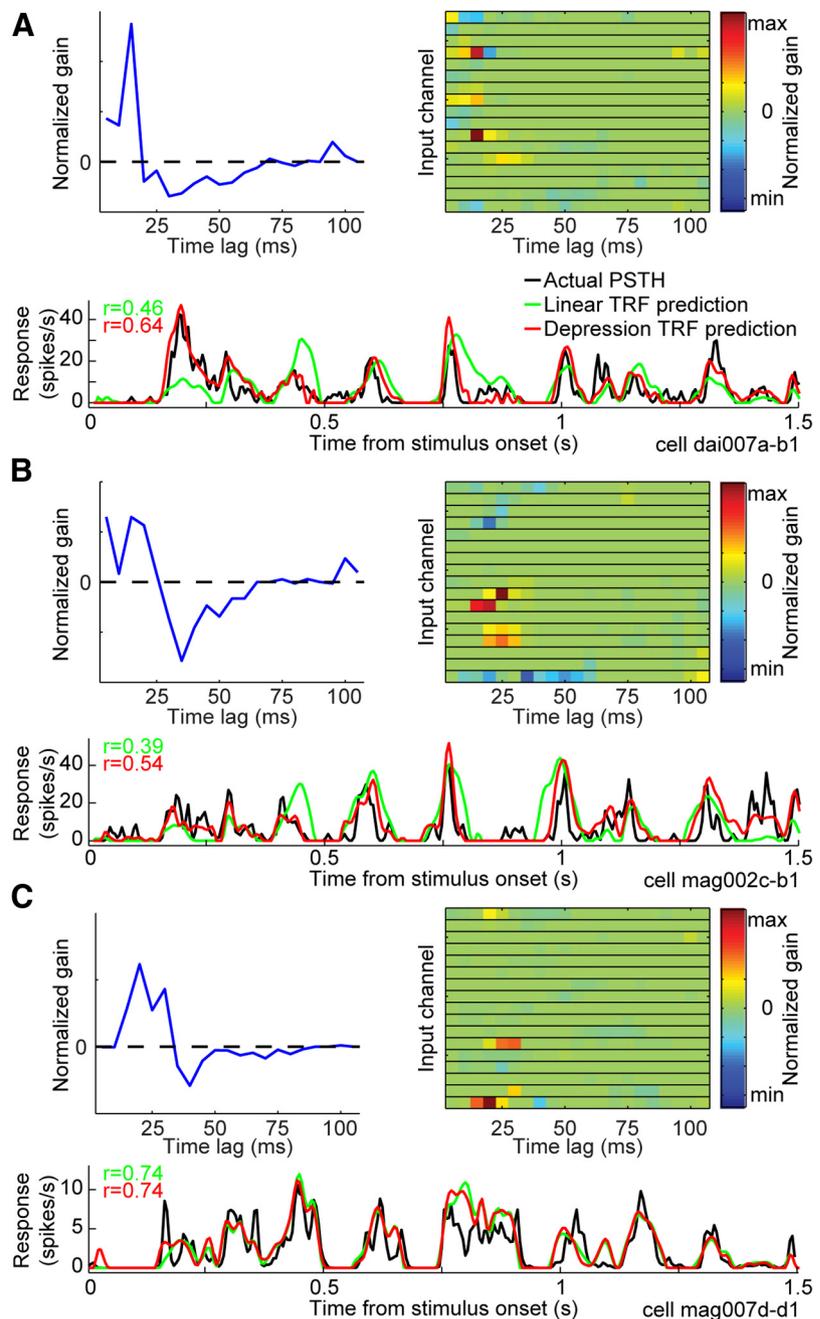


Figure 3. *A*, Example linear and depression TRFs for an A1 neuron. The linear TRF (upper left, plotted as in Fig. 2B) suggests a short latency excitatory response (5–15 ms) followed by a slower inhibitory response. The depression TRF for this neuron (upper right, plotted as in Fig. 2C) shows strong excitatory weights for two of the depressing inputs and much smaller weights for the nondepressing input (bottom row). TRF accuracy was measured by the ability to predict the PSTH response to a validation stimulus (bottom, actual PSTH in black). Prediction correlation (Pearson's r) for each TRF appears in the upper left corner of the plot (color corresponds to predicted PSTH). For this neuron, the linear TRF was able to track changes in stimulus energy (green curve), but it could not predict the relative response to each syllable event, especially the large transient at 0.2 s. The depression TRF, on the other hand, did a significantly better job predicting the relative response over time (red curve, $p < 0.05$, jackknifed t test). *B*, Second example of linear and depression TRFs, plotted as in *A*. The linear TRF again shows a strong bimodal response, with positive weights (increased firing) at very short latencies and negative weights (decreased firing) at longer latencies. The depression TRF for the same neuron shows positive weights for several of the depressing inputs with short time constants and has only small negative weights for the nondepressing input. The comparison of prediction accuracy shows that both TRFs predict the relative response strength over time, but the depression TRF more accurately captures the fine dynamics of the response to each syllable ($p < 0.05$, jackknifed t test). *C*, Third example of linear and depression TRFs. This linear TRF shows a more unimodal response, with longer latency than in *A*. The depression TRF for the same neuron shows the strongest excitatory response to the nondepressing input. Both TRFs predict the neural response with similar accuracy, indicating that this neuron shows no evidence of synaptic depression.

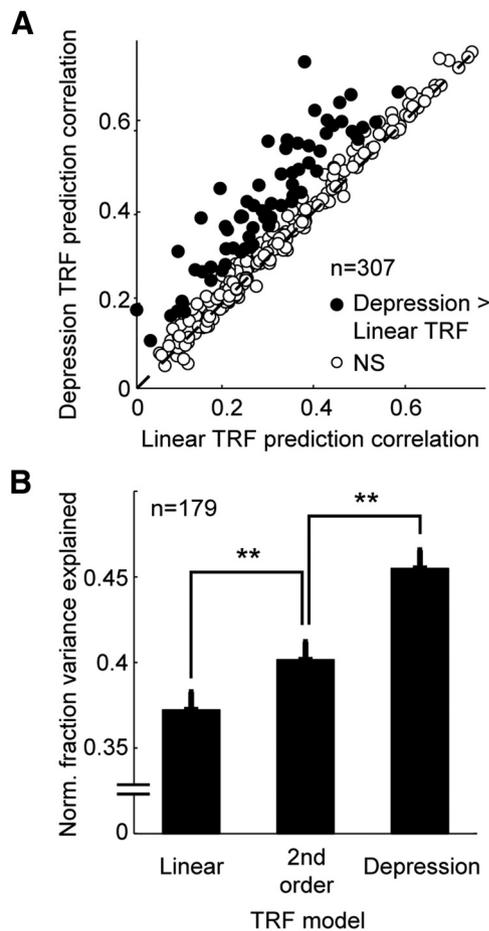


Figure 4. *A*, Comparison of prediction accuracy for linear and depression TRFs across A1 neurons ($n = 307$). Each point compares the prediction correlation for the linear TRF (horizontal axis) and depression TRF (vertical axis) estimated for a single neuron. Filled points indicate the 71 neurons for which the depression TRF prediction correlation was significantly greater than that of the linear TRF ($p < 0.05$, jackknifed t test). *B*, Mean fraction of neural response variance explained (r^2) by linear, second-order, and depression TRFs after accounting for sampling limitations in the validation dataset ($n = 179$ neurons with at least 10 repeats of the validation stimulus). Error bars indicate SEM. The second-order TRF shows an average improvement over the linear TRF (0.40 vs 0.37, $***p < 0.001$, jackknifed t test). The depression TRF performed significantly better than either other TRF (0.46, $**p < 0.0001$).

depression TRFs. Although substantially reduced from the very large short latency gain in the linear TRF, these deviations may reflect a smaller but persistent mismatch between the depression model and underlying system.

If the functional properties of the synaptic depression model capture the effects of actual underlying synaptic depression, one would expect the properties of the depressing synapses in the model to match those measured experimentally. For each synaptic depression TRF in the subset of 71 neurons showing a significant effect of depression, we measured the average time constant of depressing synapses, weighted by their relative contribution to response variance (Fig. 6A). The average time constant for these TRFs was 121 ms. Previous studies that have examined the dynamics of synaptic depression intracellularly have reported time constants on a similar, although slightly longer, time scale (Tsodyks and Markram, 1997). The histogram of time constants indicates that a large portion of A1 neurons have relatively short time constants < 100 ms, whereas a smaller number range widely from 100 to 400 ms. Although the histogram suggests a bimodal

distribution, this trend is not significant (dip test) (Hartigan and Hartigan, 1985).

Previous studies of synaptic depression have also shown that neurons with low spontaneous spike rates are more likely to show effects of synaptic depression (Hermann et al., 2007). This trend reflects the fact that input synapses with low spontaneous activity will undergo large changes in depression upon activation, whereas synapses with greater spontaneous activity will already be partially depressed before any stimulus inputs arrive. We compared spontaneous rates for neurons that showed a significant effect of depression and those that did not, and found the predicted effect. Depression neurons had significantly lower spontaneous spike rate than the nondepression group (Fig. 6B; $p < 0.001$, jackknifed t test).

As might be expected, TRF weights associated with depressing synapses were greater for neurons that showed significant effects of depression than for nondepression neurons by a factor of ~ 3 (Fig. 6C). What was not expected was that the relative weight of the nondepressing synapse was often negative for neurons that showed significant effects of depression (see also Fig. 5B). In contrast, for neurons that did not show an effect of depression, the weight of the nondepressing synapse tended to be positive. At face value, these results suggest that inhibitory inputs at the best frequency are less likely to undergo depression than excitatory inputs, but this trend toward inhibition in the nondepressing channel could also reflect nonlinear mechanisms that are not captured by the depression model.

Based on the intuition that the TRF reflects synaptic weights (Fig. 2A), the linear TRF and each row of the depression TRF can be interpreted as the postsynaptic potential evoked by an action potential arriving at a single input synapse. If a TRF estimate does reflect a synaptic potential, then the TRF weights should be unimodal across time lags. That is, the weights should be consistently either positive or negative, respectively, for excitatory or inhibitory inputs. The linear TRF often follows a pattern of excitation at short latency followed by inhibition at longer latencies (Fig. 3), which reflects the attempt of a linear temporal model to explain transient responses to stimulus onsets (Fig. 2B). Although bimodal dynamics may provide a useful functional descriptor of temporal properties, such as tuning to modulation rate (Kowalski et al., 1996), the bimodal pattern of positive-to-negative weights in the linear model likely does not reflect the input of a single synapse.

We measured an index of unimodality as the absolute value of the ratio of the average value of the TRF across time lags to the average of its absolute value. An index of 1 indicates fully unimodal weights (all positive or all negative), and an index of 0 indicates half-positive and half-negative. When we compared the unimodality index for synaptic depression and linear TRFs across neurons (Fig. 6D; for depression TRFs, the unimodality index was measured for the synapse with largest gain), the index was systematically higher for synaptic depression TRFs and the average value across neurons was significantly greater than for linear TRFs (mean depression 0.67, linear 0.28, $p < 0.0001$, randomized paired t test). Thus, the strong tendency of the depression TRF to be unimodal suggests that it provides a more biologically plausible model of synaptic potentials than the linear TRF.

Synaptic depression permits longer temporal integration

The hypothesis stated in the Introduction suggests that neurons that undergo depression should be able to encode the temporal dynamics of auditory stimuli over a longer period of time than nondepressing neurons. To measure the temporal stimulus in-

formation encoded by the neurons in our dataset, we used a nonparametric *maximum a posteriori* method for stimulus reconstruction from neural population activity. Conceptually, this method is related to linear decoding algorithms (Bialek et al., 1991; Mesgarani et al., 2009), but rather than requiring linear reconstruction filters, it uses a more general probability analysis to infer the most likely stimulus producing the neural population response. We discretized the response of each neuron based on its firing rate during the preceding 100 ms time window. Then we measured the probability of the stimulus in each preceding 5 ms bin, conditioned on the population response vector (Fig. 7A). To reconstruct the stimulus, we computed the joint stimulus probability across the neural population to infer the most probable stimulus preceding the response.

The reconstruction analysis was performed using two equal-sized groups of depression and nondepression neurons ($n = 57$), selected from the subset of neurons that had the same average prediction accuracy and that were presented stimuli with identical envelopes. Reconstruction was more accurate at longer latencies for the depression neurons than for the nondepression neurons. For depression neurons, the longest latency with reconstruction error significantly better than chance was 290 ms, whereas for nondepression neurons it was 180 ms (blue vs green curves; Fig. 7B, $p < 0.001$, jackknifed t test). Thus, after grouping neurons based on an independent criterion (performance of the linear vs depression TRF), the reconstruction analysis revealed that the two groups encoded different temporal epochs of the stimulus.

To test specifically whether synaptic depression can account for the increased temporal integration time, we repeated the reconstruction analysis, but with neural responses simulated by the linear and depression TRFs. For the depression neuron group, the depression TRFs were able to replicate nearly the same pattern of reconstruction as the actual neural data (Fig. 7C). On the other hand, linear TRFs for those same neurons were not able to reconstruct the stimulus at longer time lags. For nondepression neurons, simulated responses by both depression and linear TRFs replicated the shorter time course of accurate reconstruction (Fig. 7D). Thus, synaptic depression can explain the increased duration of stimulus encoding by the former group of neurons.

Discussion

Synaptic depression can explain the dynamics of A1 neural responses over many hundreds of milliseconds, a critical time-scale for discriminating natural sounds (Shannon et al., 1995; Mesgarani et al., 2008). When nonlinear synaptic depression is incorporated into a temporal response function (TRF), model performance improves over traditional linear TRFs, and model fits more plausibly reflect known biological properties

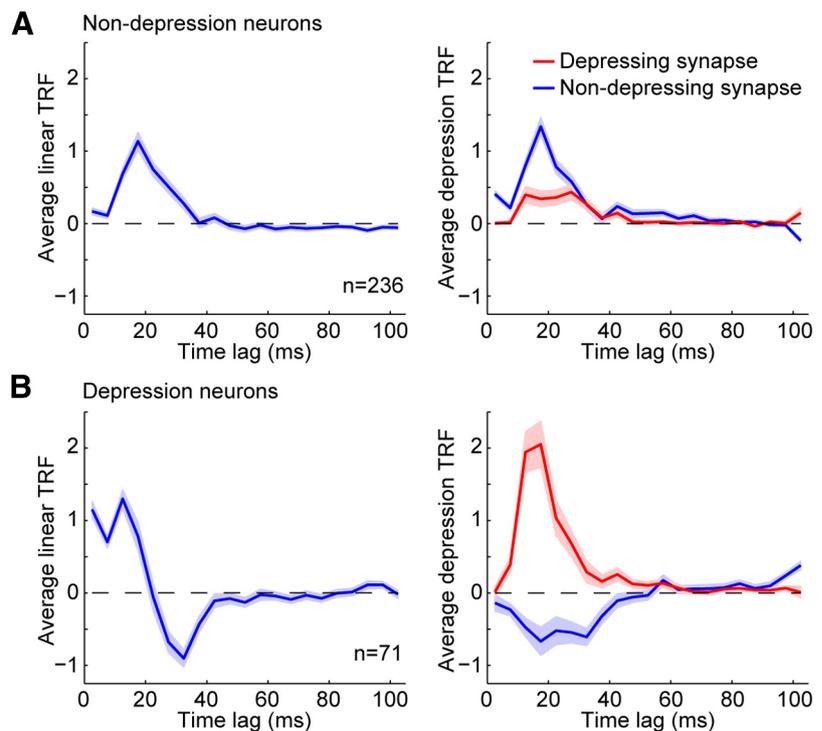


Figure 5. *A*, Average TRF properties for neurons showing no significant difference between linear and depression TRFs ($n = 236$). The average linear TRF (left) for these neurons shows a characteristically rapid excitatory response and slower decay with latency of ~ 15 ms. Shading indicates one SE across neurons after normalizing each TRF to have the same total gain. Data for the depression TRF (right) show the average temporal response for the single nondepressing synapse in blue and the average for the depressing synapse with largest gain in red. The average temporal response for the nondepressing synapse response closely resembles that of the linear TRF, and the average for the depressing synapse is much weaker. *B*, Average linear and depression TRFs for neurons that show a significant increase in predictive power for depression TRFs ($n = 71$), plotted as in *A*. For these neurons, the average linear TRF has an excitatory response with very short latency (< 5 ms) followed by an inhibitory response at longer latencies. The average depression TRF shows a more biologically plausible fit, with an excitatory response by the depressing synapse and a weaker inhibitory response by the nondepressing synapse. TRF fits were not constrained to have a particular sign or latency. The more biologically plausible latencies for the depression TRF suggest that the model better captures important temporal responses properties in A1.

of neurons. The superior performance of the depression TRF does not simply reflect its greater complexity, as a second-order Volterra model with a similar number of parameters does not perform as well. Instead, the synaptic depression model provides a better approximation of the mechanisms that actually produce responses in A1.

Synaptic depression and stimulus encoding over multiple timescales

Contextual modulation of auditory responses has been suggested to play a number of computational roles in the brain, most prominently in outlier detection (Näätänen et al., 2001; Ulanovsky et al., 2003) and adaptation to stimulus variance (Dean et al., 2008; Asari and Zador, 2009). The data reported here support the idea that synaptic depression not only contributes to these processes but also plays a more general role in encoding stimulus dynamics (Penner and Shiffrin, 1980). The envelopes of speech and other natural sounds consist of complex, irregular modulations lasting hundreds of milliseconds that contain behaviorally relevant information (Shannon et al., 1995; Singh and Theunissen, 2003; Mesgarani and Shamma, 2005). The auditory cortex must encode the stimulus over this period to guide behavior in downstream areas. Receptive field models typically reveal sensory integration lasting up to only 50–75 ms (David et al., 2009), but synaptic

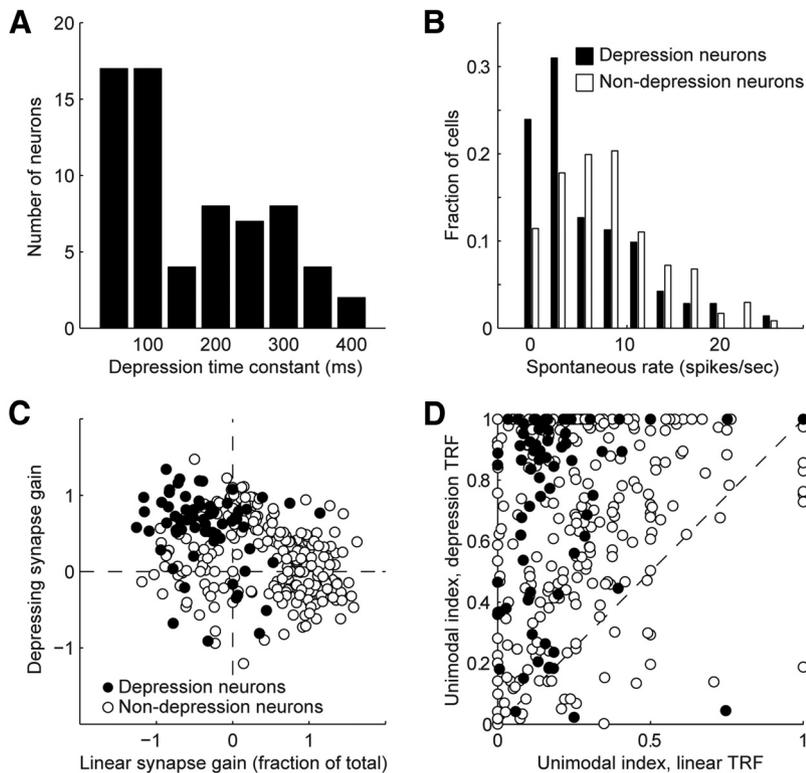


Figure 6. *A*, Histogram of depression recovery time constants for neurons showing significantly better performance by the depression TRF ($n = 71$). The average neuron recovered with a time constant of 121 ms, but values ranged widely, up to 400 ms. The trend toward a bimodal distribution was not significant (dip test). *B*, Histogram comparing average spontaneous firing rates between depression and nondepression neurons, measured during the 500 ms prestimulus epoch. Neurons showing significant evidence for depression had a lower mean firing rate (5.6 ± 0.7 spikes/s, $n = 71$) than nondepression neurons (7.9 ± 0.4 spikes/s, $n = 236$, $p < 0.001$, jackknifed t test). *C*, Comparison of gain in each depression TRF for the linear, nondepressing synapse (horizontal axis) and for the strongest depressing synapse (vertical axis). Values are the normalized fraction of the root mean square of the entire TRF. Neurons that did not show evidence for depression (open circles) tended to have positive gain on their nondepressing synapse (mean 0.45) and much weaker gain on the depressing synapse (mean 0.18). Neurons that showed evidence for depression (filled circles) often had negative gain on the nondepressing synapse (mean -0.37) and much stronger gain on the depressing synapse (mean 0.58). *D*, Comparison of unimodality of temporal responses for the linear TRF (horizontal axis) and depression TRF (vertical axis) across neurons. A unimodal index value of 1 indicates a temporal response that is entirely excitatory or inhibitory, whereas a value of 0 indicates a bimodal response that varies equally between excitatory and inhibitory across latencies. For depression TRFs, the index is plotted for the synapse (depressing or nondepressing) with highest gain. For the entire population of neurons ($n = 307$), the unimodal index is almost always greater for the depression TRF (mean 0.67) than for the linear TRF (mean 0.28, $p < 0.0001$, jackknifed t test). The difference is particularly large for depression neurons ($n = 71$, filled circles; depression TRF, mean 0.91; linear TRF, mean 0.22).

depression suggests a mechanism that can encode stimuli over these longer timescales.

Similar to previous studies using more direct experimental measurements (Tsodyks and Markram, 1997; Viana et al., 2011), we found that the dynamics of synaptic depression vary substantially across neurons. Because of this variability, each neuron encodes different contextual information, and a neural population together can encode complex sound envelope features.

Variable degrees of synaptic depression across neurons may be a general feature of cortical processing that enables the simultaneous representation of stimuli over multiple windows into the past (Buonomano and Maass, 2009). As signals travel over many synapses, integration times can accumulate and become even longer without requiring recurrent activity or long-term plasticity. This predicts that reconstruction from neural responses in downstream auditory areas should be accurate over longer time windows than in A1.

Stimulus context effects in auditory cortex

The modulatory effects of stimulus history on the responses of A1 neurons have been reported widely (Abeles and Goldstein, 1972; Brosch and Schreiner, 1997; Ulanovsky et al., 2003; Ahrens et al., 2008; Asari and Zador, 2009; Klampfl et al., 2012), but underlying mechanisms have remained largely a matter of speculation. Most previous studies used relatively small parametric stimulus sets to maximize sensitivity to modulatory effects. With limited stimulus-response data points, however, it has been difficult to constrain general receptive field models of the mechanism producing the effects. More recently, Asari and Zador (2009) studied modulation of responses to continuous natural sounds and found a variety of context effects, but the complexity of the stimulus prevented detailed modeling of mechanism. Here, we used a stimulus with intermediate complexity, having natural temporal dynamics but simple spectral structure. This hybrid natural-synthetic stimulus permitted us to determine whether a predictive filter model can explain contextual effects on neural responses to stimuli with natural temporal dynamics.

Some previously reported contextual effects last longer than the extreme of ~ 400 ms measured here (Ulanovsky et al., 2003; Asari and Zador, 2009). Whether these discrepancies reflect differences in experimental preparation (species or anesthesia levels) (Reig et al., 2006) or additional unmodeled network effects (Yaron et al., 2012) remains a question for further study.

Synaptic depression could also contribute to the mismatch negativity, an enhanced electrophysiological response to oddball stimuli observed broadly across the auditory system (Näätänen et al., 2001; Ulanovsky et al., 2003). Synapses conveying a repeated input may be sent into a state of depression. When a novel stimulus is presented that activates even a slightly different set of synapses, the nondepressed synapses will produce a stronger response that could produce the enhanced evoked potential.

Nonlinear temporal integration in A1

Although we focused on synaptic depression here, we previously considered alternative nonlinear models for dynamics in A1 during the presentation of natural stimuli (David et al., 2009). One of these models, incorporating a static spike threshold, did not account for the observed response dynamics, but it did improve model performance overall. Thus, all the models tested here incorporate a threshold nonlinearity. Another model we considered applied broadly tuned contrast gain normalization to the output of the linear filter (Carandini et al., 1997). In addition to not explaining the observed dynamics, this model was less rele-

vant to the current study, which focused on narrowband stimuli that would not activate a broadly tuned mechanism. More recent work has suggested an alternative model for gain normalization with narrow spectral tuning in A1 (Rabinowitz et al., 2012). These effects occur on a similar timescale to the depression modeled here, and further exploration of these models will determine whether they describe similar or even the same underlying mechanisms.

Nonlinear spectrotemporal models that account for second-order stimulus interactions may explain some response properties in A1, although previous studies using these models have focused on spectral rather than temporal integration (Ahrens et al., 2008; Pienkowski et al., 2009; Rabinowitz et al., 2012). The depression model resembles the second-order Volterra model, in that its computations include a squaring of the stimulus. However, the recovery time constant in the depression model reflects a basic analytical difference that permits it to have longer memory than the second-order model. Increasing maximum latency did not improve performance for either the linear or second-order TRF, given the available data. Thus, the depression TRF provides a more accurate and interpretable way to describe temporal processing in A1.

Neurons in A1 vary in their tuning for periodically modulated stimuli (Schreiner and Urbas, 1988; Eggermont, 1991), and modeling the neural population as a bank of modulation rate-tuned filters has proven useful for theoretical analysis and signal processing applications (Dau et al., 1997; Mesgarani and Shamma, 2005). Modulation tuning has been characterized phenomenologically using linear and second-order models (Kowalski et al., 1996; Depireux et al., 2012), but synaptic depression with short recovery time constants suggests a mechanism that can produce rate tuning in addition to the longer-lasting filter properties discussed above.

The synaptic depression model is a functional one, in that it does not definitively identify the mechanism controlling nonlinear processing in A1. Mechanisms other than synaptic depression, such as slow potassium currents in the postsynaptic neuron (Jolivet et al., 2004) or intracortical feedback (Yaron et al., 2012), could produce similar response dynamics. The mechanisms involved may be distinguished in future studies involving more spectrally complex stimuli or direct manipulation of neural circuits. Whatever the mechanism, however, it must engage dynamics more like the synaptic depression model than the other models tested in this study.

A model accounting for nonlinear temporal processing provides a new tool for studying the state dependence of auditory

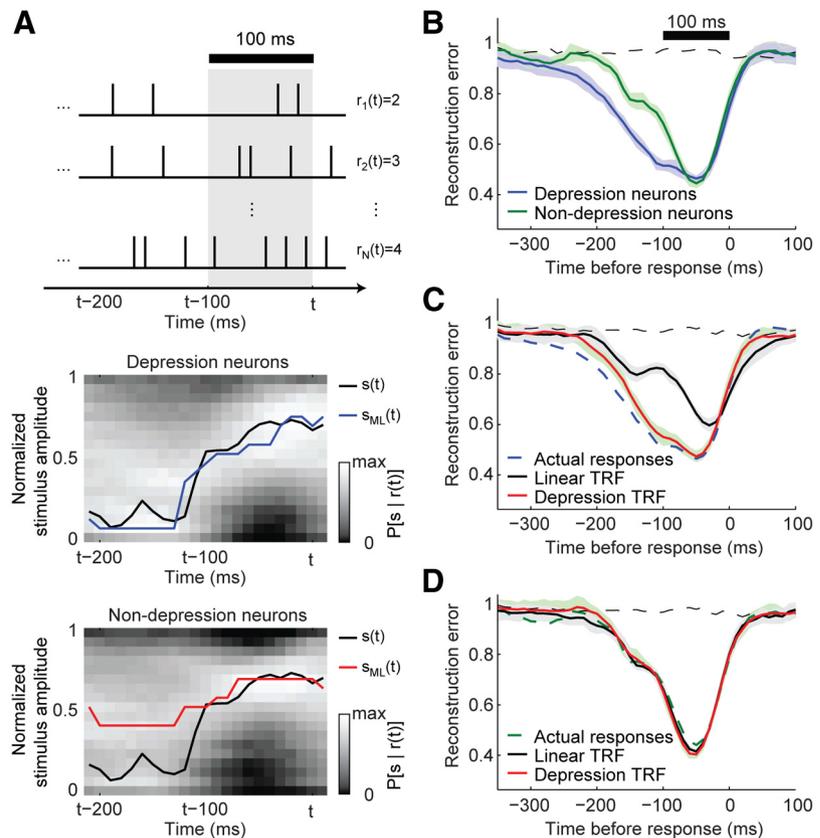


Figure 7. *A*, Schematic of the *maximum a posteriori* procedure for reconstructing the preceding stimulus envelope dynamics, based on the current neural population firing rate. At each point in time, the mean firing rate of each neuron was measured over the previous 100 ms (top). Using an estimation dataset, the probability distribution of stimulus amplitude was computed at each preceding time lag, conditioned on the neural response (heat maps in bottom panels). The stimulus was reconstructed using a separate validation dataset by taking the value at the peak of the probability distribution for the current response. Reconstruction from depression neuron responses ($n = 57$, blue line, middle) often matched the actual stimulus (black line) up to 200 ms in the past. Reconstruction from an equal number of nondepression neurons tended to be accurate only at shorter time lags (green line, bottom panel). When no information was available from the neural response, the reconstruction simply reverted to the mean stimulus level (longer time lags, bottom). *B*, Comparison of normalized mean square error for reconstruction from responses of neurons that showed evidence of depression (blue) and neurons that did not (green). Shading indicates SE, computed by jackknifing. Baseline error (dashes) was computed by randomly shifting the spike times of each neuron relative to the stimulus before reconstruction. Neurons showing evidence for depression were able to reconstruct significantly better than chance back to 290 ms ($p < 0.05$, jackknifed t test). Nondepression neurons were able to significantly reconstruct better than chance only 180 ms into the past. *C*, Reconstruction accuracy for simulated responses by TRF fits for the 57 depression neurons. Accuracy of reconstruction by simulated depression TRF responses (red) closely matched reconstruction from the actual responses (blue dashes, replotted from *B*). Reconstruction from simulated responses by linear TRFs for the same neurons (black) did not reconstruct the stimulus at longer time lags. *D*, Reconstruction accuracy for simulated responses by TRF fits for nondepression neurons. The simulated responses of both depression and linear TRFs were able to reconstruct the stimulus over the same latency range as the actual neural data (green dashes, replotted from *B*).

representations. The influence of brain state on sensory responses in A1 is well established, reflecting both explicit behavioral control (David et al., 2012) and uncontrolled changes in network state (Marguet and Harris, 2011). Any effects of internal state on synaptic depression or depression-like mechanisms can be observed explicitly in models fit in these different conditions.

Inferring biological circuits from computational models

In his classic work on the visual system, Marr (1982) described three approaches to characterizing sensory processing: computational, algorithmic, and implementational. Since then, most new spectrotemporal models have been motivated largely at computational (Atencio et al., 2008; Pienkowski et al., 2009) and algorithmic levels (Ahrens et al., 2008; Rabinowitz et al., 2012). Here, we took a less common approach, motivated at the implementa-

tional level (but see Gill et al., 2006; Ozuysal and Baccus, 2012; Schinkel-Bielefeld et al., 2012). Our model incorporated synaptic depression, a nonlinear mechanism known to operate in the cortex, and proved more effective than the computationally motivated second-order Volterra model. Nonlinear models linked to physiological mechanisms are widely accepted for the sensory periphery (Lyon and Mead, 1988; Yang et al., 1992), and our findings suggest that the benefits of mechanism-based models extend more centrally. Perhaps more importantly, the depression model points to a mechanism that can explain its improved performance, and future studies can test whether the putative depression mechanism does indeed produce the observed functional effects.

The enduring value of a new model for sensory processing can be tested in the real world. Important concepts that have emerged from spectrotemporal models of the auditory system have been used to improve algorithms for speech and sound processing (Mesgarani and Shamma, 2005). The results reported here demonstrate that algorithms mimicking synaptic depression can encode complex stimulus dynamics over a longer time period than a traditional linear model. Further investigation will reveal whether new algorithms derived from biological circuits provide better solutions to existing signal processing problems.

References

- Abeles M, Goldstein MH Jr (1972) Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. *Brain Res* 42:337–352. [CrossRef Medline](#)
- Ahrens MB, Linden JF, Sahani M (2008) Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci* 28:1929–1942. [CrossRef Medline](#)
- Asari H, Zador AM (2009) Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *J Neurophysiol* 102:2638–2656. [CrossRef Medline](#)
- Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58:956–966. [CrossRef Medline](#)
- Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252:1854–1857. [CrossRef Medline](#)
- Bizley JK, Nodal FR, Nelken I, King AJ (2005) Functional organization of ferret auditory cortex. *Cereb Cortex* 15:1637–1653. [CrossRef Medline](#)
- Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702. [CrossRef Medline](#)
- Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* 77:923–943. [Medline](#)
- Buonomano DV, Maass W (2009) State-dependent computations: spatio-temporal processing in cortical networks. *Nat Rev Neurosci* 10:113–125. [CrossRef Medline](#)
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644. [Medline](#)
- Carandini M, Heeger DJ, Senn W (2002) A synaptic explanation of suppression in visual cortex. *J Neurosci* 22:10053–10065. [Medline](#)
- Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* 18:4785–4799. [Medline](#)
- Dau T, Kollmeier B, Kohlrausch A (1997) Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow-band carriers. *J Acoust Soc Am* 102:2892–2905. [CrossRef Medline](#)
- David SV, Gallant JL (2005) Predicting neuronal responses during natural vision. *Network* 16:239–260. [CrossRef Medline](#)
- David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectrotemporal receptive fields with natural stimuli. *Network* 18:191–212. [CrossRef Medline](#)
- David SV, Mesgarani N, Fritz JB, Shamma SA (2009) Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *J Neurosci* 29:3374–3386. [CrossRef Medline](#)
- David SV, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A* 109:2144–2149. [CrossRef Medline](#)
- Dean I, Robinson BL, Harper NS, McAlpine D (2008) Rapid neural adaptation to sound level statistics. *J Neurosci* 28:6430–6438. [CrossRef Medline](#)
- deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443. [CrossRef Medline](#)
- Depireux DA, Dobbins HD, Marvit P, Shechter B (2012) Dynamics of phase-independent spectro-temporal tuning in primary auditory cortex of the awake ferret. *Neuroscience* 214:28–35. [CrossRef Medline](#)
- Eggermont JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. *Hearing Res* 56:153–167. [CrossRef Medline](#)
- Eggermont JJ (1993) Wiener and Volterra analysis applied to the auditory system. *Hearing Res* 66:177–201. [CrossRef Medline](#)
- Eichhorn J, Tolias A, Zien A, Kuss M (2004) Prediction on spike data using kernel algorithms. In: *Advances in neural information processing systems*, Vol 16 (Becker S, Thrun S, Obermayer K, eds), pp 1367–1374. Cambridge, MA: MIT.
- Garcia-Lazaro JA, Ahmed B, Schnupp JW (2006) Tuning to natural stimulus dynamics in primary auditory cortex. *Curr Biol* 16:264–271. [CrossRef Medline](#)
- Gill P, Zhang J, Woolley SM, Fremouw T, Theunissen FE (2006) Sound representation methods for spectro-temporal receptive field estimation. *J Comput Neurosci* 21:5–20. [CrossRef Medline](#)
- Hartigan JA, Hartigan PM (1985) The dip test of unimodality. *Ann Statist* 13:70–84. [CrossRef](#)
- Hermann J, Pecka M, von Gersdorff H, Grothe B, Klug A (2007) Synaptic transmission at the calyx of Held under in vivo like activity levels. *J Neurophysiol* 98:807–820. [CrossRef Medline](#)
- Jolivet R, Lewis TJ, Gerstner W (2004) Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *J Neurophysiol* 92:959–976. [CrossRef Medline](#)
- Klampfl S, David SV, Yin P, Shamma SA, Maass W (2012) A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J Neurophysiol* 108:1366–1380. [CrossRef Medline](#)
- Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9:85–111. [CrossRef Medline](#)
- Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76:3503–3523. [Medline](#)
- Lyon RF, Mead C (1988) An analog electronic cochlea. *IEEE Trans Acoustics Speech Signal Processing* 36:1119–1134. [CrossRef](#)
- Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24:1089–1100. [CrossRef Medline](#)
- Marguet SL, Harris KD (2011) State-dependent representation of amplitude-modulated noise stimuli in rat auditory cortex. *J Neurosci* 31:6414–6420. [CrossRef Medline](#)
- Marr D (1982) *Vision*. San Francisco: Freeman.
- Mesgarani N, Shamma SA (2005) Speech enhancement based on filtering the spectrotemporal modulations. *IEEE Int Conf Acoustics Speech Signal Processing* 1105–1108.
- Mesgarani N, David SV, Fritz JB, Shamma SA (2008) Phoneme representation and classification in primary auditory cortex. *J Acoust Soc Am* 123:899–909. [CrossRef Medline](#)
- Mesgarani N, David SV, Fritz JB, Shamma SA (2009) Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J Neurophysiol* 102:3329–3339. [CrossRef Medline](#)
- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) “Primitive intelligence” in the auditory cortex. *Trends Neurosci* 24:283–288. [CrossRef Medline](#)
- Ozuysal Y, Baccus SA (2012) Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron* 73:1002–1015. [CrossRef Medline](#)
- Penner MJ, Shiffrin RM, Shiffrin RM (1980) Nonlinearities in the coding of intensity within the context of a temporal summation model. *J Acoust Soc Am* 67:617–627. [CrossRef Medline](#)
- Pienkowski M, Shaw G, Eggermont JJ (2009) Wiener-Volterra characterization of neurons in primary auditory cortex using Poisson-distributed impulse train inputs. *J Neurophysiol* 101:3031–3041. [CrossRef Medline](#)

- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2012) Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J Neurosci* 32:11271–11284. [CrossRef Medline](#)
- Reig R, Gallego R, Nowak LG, Sanchez-Vives MV (2006) Impact of cortical network activity on short-term synaptic depression. *Cereb Cortex* 16:688–695. [CrossRef Medline](#)
- Sahani M, Linden JF (2003) How linear are auditory cortical responses? In: *Advances in neural information processing systems* (Becker S, Thrun S, Obermayer K, eds), pp 301–308. Cambridge, MA: MIT.
- Schinkel-Bielefeld N, David SV, Shamma SA, Butts DA (2012) Inferring the role of inhibition in auditory processing of complex natural stimuli. *J Neurophysiol* 107:3296–3307. [CrossRef Medline](#)
- Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat: II. Comparison between cortical fields. *Hearing Res* 32:49–63. [CrossRef Medline](#)
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304. [CrossRef Medline](#)
- Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114:3394–3411. [CrossRef Medline](#)
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatial temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316. [CrossRef Medline](#)
- Tsodyks M, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci U S A* 94:719–723. [CrossRef Medline](#)
- Tsodyks M, Pawelzik K, Markram H (1998) Neural networks with dynamic synapses. *Neural Comput* 10:821–835. [CrossRef Medline](#)
- Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391–398. [CrossRef Medline](#)
- Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. *J Neurosci* 24:10440–10453. [CrossRef Medline](#)
- Viaene AN, Petrof I, Sherman SM (2011) Synaptic properties of thalamic input to layers 2/3 and 4 of primary somatosensory and auditory cortices. *J Neurophysiol* 105:279–292. [CrossRef Medline](#)
- Won JH, Drennan WR, Nie K, Jameyson EM, Rubinstein JT (2011) Acoustic temporal modulation detection and speech perception in cochlear implant listeners. *J Acoust Soc Am* 130:376–388. [CrossRef Medline](#)
- Wright BA, Bowen RW, Zecker SG (2000) Nonlinguistic perceptual deficits associated with reading and language disorders. *Curr Opin Neurobiol* 10:482–486. [CrossRef Medline](#)
- Wu MC, David SV, Gallant JL (2006) Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29:477–505. [CrossRef Medline](#)
- Yang X, Wang K, Shamma SA (1992) Auditory representations of acoustic signals. *IEEE Trans Info Theory* 38:824–839. [CrossRef](#)
- Yaron A, Hershenhoren I, Nelken I (2012) Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron* 76:603–615. [CrossRef Medline](#)