Complementary effects of adaptation and gain control on sound encoding in primary auditory cortex

Jacob Pennington[1] and Stephen David[*2]

[1]Department of Mathematics, Washington State University Vancouver
[2]Oregon Hearing Research Center, Oregon Health and Science University

January 15, 2020

Abstract

A common model for the function of auditory cortical neurons is the linear-nonlinear spectro-temporal receptive field (LN STRF). This model casts the neural spike rate at each moment as a linear weighted sum of the preceding sound spectrogram, followed by nonlinear rectification. While the LN model can account for many aspects of auditory coding, it fails to account for long-lasting effects of sensory context on sound-evoked activity. Two models have expanded on the LN STRF to account for these contextual effects, using short-term plasticity (STP) or contrast-dependent gain control (GC). Both models improve performance over the LN model, but they have never been compared directly. Thus, it is unclear whether they account for distinct processes or describe the same phenomenon in different ways. To address this question, we recorded activity of primary auditory cortical neurons in awake ferrets during presentation of natural sound stimuli. We fit models incorporating one nonlinear mechanism (GC or STP) or both (GC+STP) on this single dataset. We compared model performance according to prediction accuracy on a held-out dataset not used for fitting. The STP model performed significantly better than the GC model, but the GC+STP model performed significantly better than either individual model. We also quantified equivalence between the STP and GC models by calculating the partial correlation between their predictions, relative to the LN model. We found only a modest degree of equivalence between them. We observed similar results for a smaller dataset collected in clean and noisy acoustic contexts. Together, the improved performance of the combined model and weak equivalence between STP and GC models suggest that they describe distinct processes. Models incorporating both mechanisms are necessary to fully describe auditory cortical coding.

*Correspondence: davids@ohsu.edu

Introduction

The evoked activity of an auditory neuron can be modeled as a function of a time-varying stimulus plus some amount of error, reflecting activity that cannot be explained by the model. Sources of error can include physiological noise, recording artifacts, and model misspecification. A common encoding model used to study the auditory system is the linear-nonlinear spectro-temporal receptive field (LN) model (Calabrese et al. 2011; David 2018; Rahman et al. 2019). Accoring to the LN model, the time-varying response of a neuron can be predicted by convolution of a linear filter with the sound spectrogram followed by a static rectifying nonlinearity. The LN model has been useful because of its generality; that is, because it provides a representation of a neuron's response properties that holds for arbitrary stimuli (Aertsen and Johannesma 1981).

Despite the relative value of the LN model, it fails to account for important aspects of auditory coding. Auditory neurons often respond differently to identical sounds based on sensory context (Hiroki and Zador 2009; Rabinowitz, Willmore, Schnupp, et al. 2012). These contextual effects can be long-lasting and reflect fundamentally nonlinear computations outside of the scope of a linear filter (David and Shamma 2013). Several studies have attempted to bridge this gap in modeling capacity by extending the LN model to incorporate experimentally-observed biological phenomena like short-term synaptic plasticity (STP) and contrast-dependent gain control (GC) (Cooke et al. 2018; Lopez Espejo, Schwartz, and David 2019; Rabinowitz, Willmore, Schnupp, et al. 2012). These models show improved performance over the LN model, and point to mechanisms that explain contextual effects.

The success of these alternatives to the LN model is well established, but the extent to which they describe distinct, complementary mechanisms within the brain is unclear. It has been suggested, for example, that short-term plasticity may in fact underlie contrast-dependent gain control (Carandini, Heeger, and Senn 2002; Rabinowitz, Willmore, Schnupp, et al. 2011). At the same time, both gain control and STP have been implicated in the robust coding of natural stimuli in noise (Mesgarani et al. 2014; Rabinowitz, Willmore, King, et al. 2013).

The complementarity of these effects has been difficult to test because their respective models were implemented separately and tested on different datasets. To address this issue, we tested both STP and GC models using identical natural sound datasets, collected from ferret primary auditory cortex (A1). The first set was comprised of a large collection of natural sounds. The second was a set of ferret vocalizations with and without additive broadband noise. We focused on natural sound coding because LN models fit using synthetic stimuli do not do a good job of predicting responses to natural sounds presented to the same neurons (David, Mesgarani, et al. 2009).

With both models on equal footing, we compared their performance to a classic LN model and to each other. Both models showed improved performance over the LN model, but a model combining the STP and GC mechanisms performed better than either one alone.

Additionally, we found a low degree of commonality between the STP and GC models' predictions after accounting for the LN model's contributions. These results suggest that models for short-term plasticity and contrast-dependent gain control are not equivalent, but in fact offer complementary explanations of auditory cortical coding.

Materials and Methods

Experimental Procedures

Data collection. All procedures were approved by the University Institutional Animal Care and Use Committee and conform to standards of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Prior to experiments, animals were implanted with a custom 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 (Synthes, 6mm) were attached to the skull, the head post was glued on the mid-line (3M Durelon), and the site was covered with bone cement (Zimmer Palacos). After surgery, the skin around the implant was allowed to heal. Analgesics and antibiotics were administered under veterinary supervision until recovery.

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. Neurophysiological activity was recorded using tungsten microelectrodes (1–5 MO, A.M. Systems). One to four electrodes positioned by independent microdrives (Alpha-Omega Engineering EPS) were inserted into the cortex.

Electrophysiological activity was amplified (A.M. Systems 3600), digitized (National Instruments PCI-6259), and recorded using the MANTA open-source data acquisition software (Englitz et al. 2013). Recording site locations were confirmed as being in A1 based on tonotopy, relatively well-defined frequency tuning and short response latency (Kowalski, Depireux, and Shamma 1996).

Spiking events were extracted from the continuous raw electrophysiological trace by principal components analysis and k-means clustering (David, Mesgarani, et al. 2009). 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 units with greater than 80% isolation were used for analysis.

Stimulus presentation was controlled by custom software written in Matlab (version 2012A, Mathworks). Digital acoustic signals were transformed to analog (National Instruments PCI6259) and amplified (Crown D-75a) to the desired sound level. Stimuli were presented

through a flat-gain, free-field speaker (Manger) 80 cm distant, 0-deg elevation and 30-deg azimuth contralateral to the neurophysiological recording site. Prior to experiments, sound level was calibrated to a standard reference (Brüel & Kjær). Stimuli were presented at 60–65 dB SPL.

Natural stimuli. The majority of data included in this study were collected during presentation of a library of 93, 3-sec natural sounds (Lopez Espejo, Schwartz, and David 2019). A larger number of these sounds (33/93) were ferret vocalizations. The vocalizations were recorded in a sound-attenuating chamber using a commercial digital recorder (44-KHz sampling, Tascam DR-400). Recordings included infant calls (1 week to 1 month of age), adult aggression calls, and adult play calls. No animals that produced the vocalizations in the stimulus library were used in the current study. The remaining natural sounds were drawn from a large library of human speech, music and environmental noises developed to characterize natural sound statistics (McDermott, Schemitsch, and Simoncelli 2013).

Neural activity was recorded during 3 repetitions of 90 of these stimuli in random order and 24 repetitions of the remaining 3 (all ferret vocalizations), presented on random interleaved trials. The low-repetition data were used for model estimation and the high-repetition data were used for model validation.

A second dataset was collected during presentation of ferret vocalizations in clean and noisy conditions. Forty 3-sec vocalizations were each presented without distortion (clean) and with additive Gaussian white noise (0 dB SNR, peak-to-peak). The noise started 0.5 sec prior to the onset and ended 0.5 sec following the offset of each vocalization. A distinct frozen noise sample was paired with each vocalization to allow repetition of identical noisy stimuli. Stimuli were presented at 65 dB SPL with 1-sec inter-stimulus interval.

Modeling framework.

Cochlear filterbank To represent the input for all the encoding models, stimulus waveforms were converted into spectrograms using a second-order gammatone filterbank (Katsiamis, Drakakis, and Lyon 2007). The filterbank included $C = 18$ filters with f_i spaced logarithmically from $f_{low} = 200$ to $f_{high} = 20,000$ Hz. After filtering, the signal was smoothed and down-sampled to 100 Hz to match the temporal bin size of the PSTH, and log compression was applied to account for the action of the cochlea (Thorson, Liénard, and David 2015).

Linear-nonlinear spectrotemporal receptive field (LN) model. The first stage of the LN STRF applied a finite impulse response (FIR) filter, h, to the stimulus spectrogram, s, to generate a linear firing rate prediction (y_{lin}):

$$y_{lin}(t) = \sum_{f}^{F} \sum_{u}^{U} h_{f,u}\, s(f, t-u)$$

For this study, the filter consisted of $F = 18$ spectral channels and $U = 15$ temporal bins (10ms each). In principle, this step can be applied to the spectrogram as a single 18x15 filter. In practice, the filter was applied in two stages: multiplication by an 18x3 spectral weighting matrix followed by convolution with a 3x15 temporal filter. Previous work has shown that this strategy is advantageous (Thorson, Liénard, and David 2015).

The output of the filtering operation was then used as the input to a static sigmoid nonlinearity that mimicked spike threshold and firing rate saturation to produce the final model prediction. For this study, we used a double exponential nonlinearity (Thorson, Liénard, and David 2015):

$$y(t) = b + a\, e^{-e^{-e^{k}(y_{lin}(t)-s))}}$$

where the baseline spike rate, saturated firing rate, firing threshold, and gain are represented by the parameters b, a, s, and k, respectively.

Short-term plasticity. The output of each spectral channel served as the input to a virtual synapse that could undergo either depression or facilitation (Lopez Espejo, Schwartz, and David 2019; Thorson, Liénard, and David 2015; Tsodyks, Pawelzik, and Markram 1998). In this model, the number of presynaptic vesicles available for release is dictated by the fraction of vesicles released by previous stimulation, u, and a recovery time constant, τ. For depression, $u > 0$, and the fraction of available vesicles, $d(t)$, is updated,

$d_i(t) = d_i(t-1) - us_i(t-1)d_i(t-1) + \frac{1-d_i(t-1)}{\tau}$

For facilitation, $u < 0$, and $d(t)$ is updated,

$d_i(t) = d_i(t-1) - us_i(t-1)[2 - d_i(t-1)] + \frac{1-d_i(t-1)}{\tau}$

The input to the i-th synapse, s_i, is scaled by the fraction of available vesicles, d_i:

$s_{di}(t) = d_i(t)s_i(t)$

The scaled output, $s_{di}(t)$, is then used as the input to the temporal filtering stage of the LN STRF.

Contrast-dependent gain control. The contrast-dependent gain control model was adapted from Rabinowitz et al. 2012. In this model, the parameters of the output nonlinearity depend on a linear weighting of the time-varying stimulus contrast. For each stimulus, the contrast, C, within a frequency band,f, was calculated as the coefficient of variation,

$$C_f(t) = \frac{\sigma_f(t)}{\mu_f(t)}$$

within a 70ms rolling window offset by 20ms, where σ_f is the standard deviation and μ_f is the mean level (dB SPL). In this model's original formulation, a linear filter with fittable coefficients would then be applied to the stimulus contrast. For this study, we simplified the model by assuming a fixed filter that simply summed stimulus contrast across frequencies, without any temporal effects. The output, K, of this summation was then used to determine the parameters of the output nonlinearity such that the i-th parameter, θ_i, was determined from a base value, θ_{i_0}, and a contrast weighting term, θ_{i_1}:

$$K(t) = \sum_{f}^{F} C_f(t)$$

$$\theta_i(t) = \theta_{i_0} + (\theta_{i_1} - \theta_{i_0})K(t)$$

Model optimization. Models were optimized using an L-BFGS-B gradient descent algorithm from the SciPy Python library. This procedure was used to minimize the mean-squared error (MSE) between a neuron's time-varying response, averaged across any repeated presentations of the same stimulus, and the model's prediction. A shrinkage correction was applied to the MSE measurement to provide an effective, flexible early stopping criterion (Thorson, Liénard, and David 2015). Post-fitting performance was evaluated based on the correlation coefficient (Pearson's R) between prediction and response, adjusted for the finite sampling of validation data (Hsu, Borst, and Theunissen 2004).

Equivalence analysis. Equivalence of STP and GC models was quantified using the partial correlation between the PSTH response to the validation stimuli predicted by each model, computed relative to the prediction by the LN model. If both models differed from the LN model in exactly the same way, the partial correlation would be 1.0. If they deviated in completely different ways, the partial correlation would be 0.0.

Results

Models incorporating short-term plasticity and contrast-dependent gain control in neural encoding of natural sounds

Previous studies have argued separately that both short-term synaptic plasticity (STP) and contrast-dependent gain control (GC) can influence the encoding of complex sounds

in primary auditory cortex (A1) (Lopez Espejo, Schwartz, and David 2019; Rabinowitz, Willmore, Schnupp, et al. 2011; Wehr and Zador 2005). Encoding models incorporating either mechanism predict time-varying neural activity in A1 better than classical linear-nonlinear spectro-temporal (LN) models. However, these models have never been compared directly using a single data set. Thus, it remains unclear whether they account for the same or complementary aspects of neural coding. To compare these nonlinearities, we recorded the activity of $n = 540$ neurons in A1 of awake, non-behaving ferrets during presentation of a large natural sound library (Fig. 1, Lopez Espejo, Schwartz, and David 2019). We then used an encoding model approach to compare how the STP and GC nonlinearities accounted for sound-evoked activity.

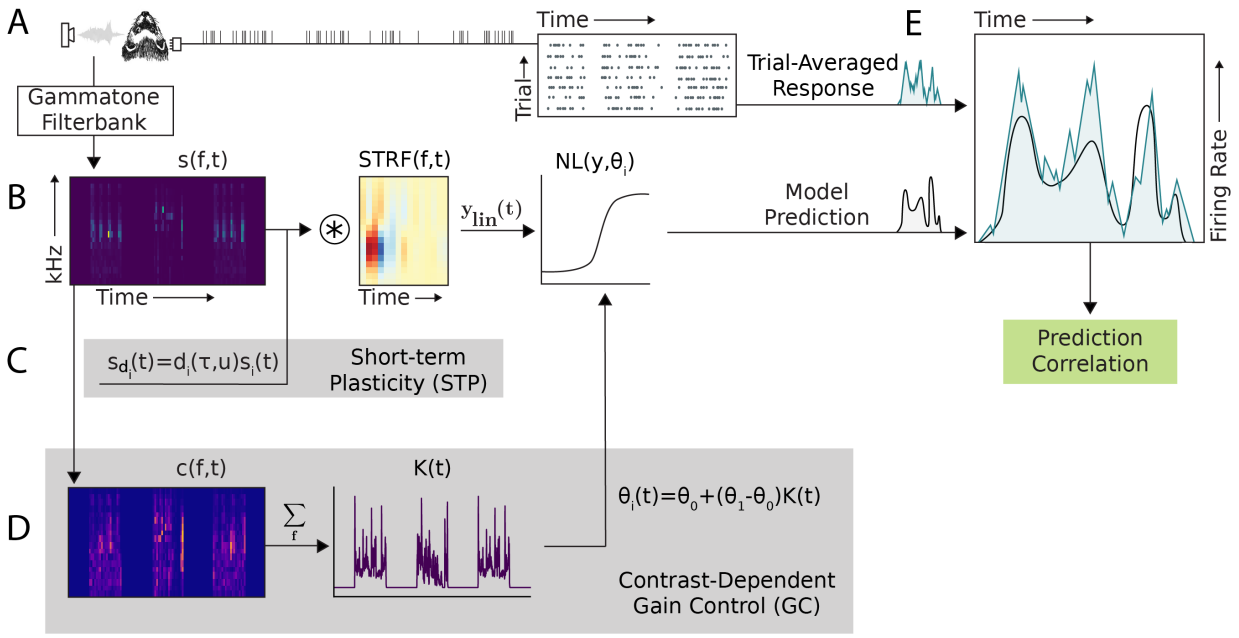

Figure 1: Schematic of four different architectures for modeling sound encoding by neurons in auditory cortex. (A) Single neuron activity was recorded from primary auditory cortex (A1) of awake, passively listening ferrets during presentation of a large set of natural sound stimuli. The trial-averaged response to each sound was calculated as the instantaneous firing rate using 10ms bins. Sound waveforms were transformed into 18-channel spectrograms with log-spaced frequencies for input to the models. (B) Linear nonlinear spectro-temporal receptive field (LN) model: stimulus spectrogram is convolved with a fitted STRF followed by nonlinear rectification. (C) Short-term plasticity (STP) model: simulated synapses depress or facilitate spectral stimulus channels prior to temporal convolution. (D) Contrast-dependent gain control (GC) model: the coefficient of variation (contrast) of the stimulus spectrogram within a rolling window is summed across frequencies. Parameters for the nonlinear rectifier are scaled by the time-varying contrast. (E) Model performance is measured using the correlation coefficient (Pearson's R) between the trial-averaged response and the model prediction. The four architectures were defined as follows - LN: B only, STP: B and C, GC: B and D, GC+STP: B, C, and D.

We compared performance of four model architectures, fitting and evaluating each with the same data set (Fig. 1b-d). The first was a standard LN model, which is widely used to characterize spectro-temporal sound encoding properties (Calabrese et al. 2011; Rahman

et al. 2019; Simoncelli et al. 2003) and provided a baseline for the current study. The second architecture (STP model) accounted for synaptic depression or facilitation by scaling input stimuli through simulated plastic synapses (Tsodyks, Pawelzik, and Markram 1998; Wehr and Zador 2005). The third (GC model) mimicked shifts in a neuron's sound-evoked spike rate as a function of recent stimulus contrast (Rabinowitz, Willmore, Schnupp, et al. 2011). A fourth architecture (GG+STP model) added both STP and GC mechanisms to the LN model. The LN, STP and GC models were implemented following previously published architectures Lopez Espejo, Schwartz, and David 2019; Rabinowitz, Willmore, Schnupp, et al. 2012, and the GC+STP model combined elements from the other models in a single architecture. The same subset of data from each neuron was used to fit the models, and performance was evaluated using a reserved validation set that was not used for fitting (see Methods). Model fitting and testing was performed using the NEMS Python library (https://github.com/LBHB/nems).

Complementary explanatory power by short-term plasticity and gain control models.

We quantified prediction accuracy using the correlation coefficient (Pearson's R) between the predicted and actual peri-stimulus time histogram (PSTH) response in each neuron's validation data. Before comparing performance between models, we identified auditory-responsive neurons for which the prediction accuracy of all of the four models was greater than expected for a random prediction ($p < 0.05$, permutation test, $n = 468/540$). Comparisons then focused on this subset. This conservative choice ensured that comparisons between model performance were not disproportionately impacted by cells for which a particular model's optimization failed.

We then compared average prediction correlations across the significantly auditory neurons. The statistical significance of differences in median prediction correlation were determined via two-sided Wilcoxon signed-rank tests.

A comparison of median prediction correlations ($n = 468$, Fig. 2a) revealed that both the GC model and the STP model performed significantly better than the LN model ($p = 7.54 \times 10^{-30}$ and $p = 3.55 \times 10^{-26}$, respectively), which confirmed previous results (Lopez Espejo, Schwartz, and David 2019; Rabinowitz, Willmore, Schnupp, et al. 2012). We also found that the STP model performed significantly better than the GC model ($p = 0.0005$). If the STP and GC models were equivalent to one another, we would not expect to observe further improvement for the combined GC+STP model. Instead, we observed a significant increase in predictive power for the combined model compared to both the GC model and the STP model ($p = 5.10 \times 10^{-}23$ and $p = 2.02 \times 10^{-}27$, respectively). The improvement for the combined model suggests that the STP and GC models describe complementary functional properties of A1 neurons.

The scatter plot in Figure 2b compares performance of the LN and combined models for

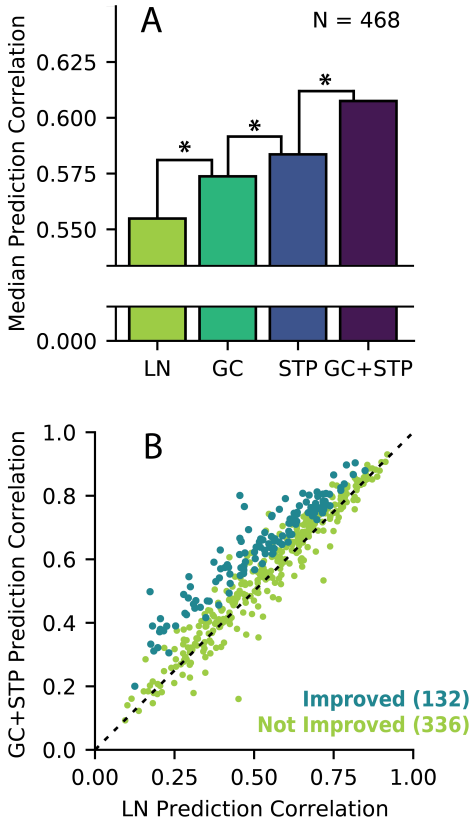

Figure 2: Comparison of prediction accuracy for each model. (A) Median prediction correlation for each model ($n = 468$ neurons). Differences were all significant between LN and GC ($p = 7.54 \times 10^{-30}$), GC and STP ($p = 0.0005$), and between STP and GC+STP models ($p = 2.02 \times 10^{-27}$, * ¡ 0.05, Wilcoxon signed-rank test). (B) Scatter plot compares prediction correlations for the LN model (horizontal axis) versus the combined GC+STP model (vertical axis) for each neuron, colored by whether the combined model showed a significant improvement (blue, $p < 0.05$, permutation test) or not (green).

each neuron. Among the 468 auditory-responsive neurons, 132 (28.2%) showed a significant improvement in prediction accuracy for the combined versus the LN model ($p < 0.05$, jackknife t-test). For the analyses of model equivalence and parameter distributions below, we focus on this set of improved neurons.

Limited equivalence of short-term plasticity and contrast gain models.

A central question in this study was the extent to which the STP model's improved performance over the LN model could be accounted for by the GC model, or vice-versa. Among non-improved cells, the two models' predictions were often closely matched to each other and to the prediction of the LN model (Fig. 3a). However, for some improved neurons, the predictions by the STP and GC models were readily distinguishable not only from the LN model but also from each other (Fig. 4b,c). In this case, the models may both improve prediction accuracy, but they do so with low equivalence. That is, the models' predicted responses deviate from that of the LN model in different ways. If the STP and GC models accounted for equivalent nonlinear properties, their predicted responses should remain similar to each other even when differing from the LN model.

To quantify model equivalence across all neurons, we first compared the change in prediction correlation for the STP and GC models, relative to the LN model (Fig. 4a). If the two models were equivalent, we would expect a strong positive correlation between the models' improvements over the LN model. However, we observed only a weak correlation (r = 0.18, $p = 6.42 \times 10^{-5}$). Thus, despite some overlap, there were substantial differences between the group of neurons with activity better explained by the STP model and those better explained by the GC model.

Next we wanted to directly measure the similarity of the STP and GC models' predictions. Since both models are extensions of the LN model, this required discounting the contributions of the LN model to the predictions (Fig. 4b). We defined model equivalence for each neuron as the partial correlation between the STP and GC model predictions relative to the LN model prediction. An equivalence score of 1.0 would indicate perfectly equivalent STP and GC model predictions, and a score of 0 would indicate that the models accounted for completely different response properties.

For the STP and GC models fit to the $n = 132$ neurons with improvements over the LN model, the distribution of partial correlations had a median of 0.29. This value was relatively low, again indicating only weak equivalence between the models. However, it was significantly greater than for non-improved neurons (median 0.17) ($p = 1.72 \times 10^{-8}$, Mann-Whitney U test). This suggests that, despite the models' differences, there are some similarities between the neural dynamics accounted for by the STP and GC models.

There was still the possibility that equivalence scores were high predominantly for STP

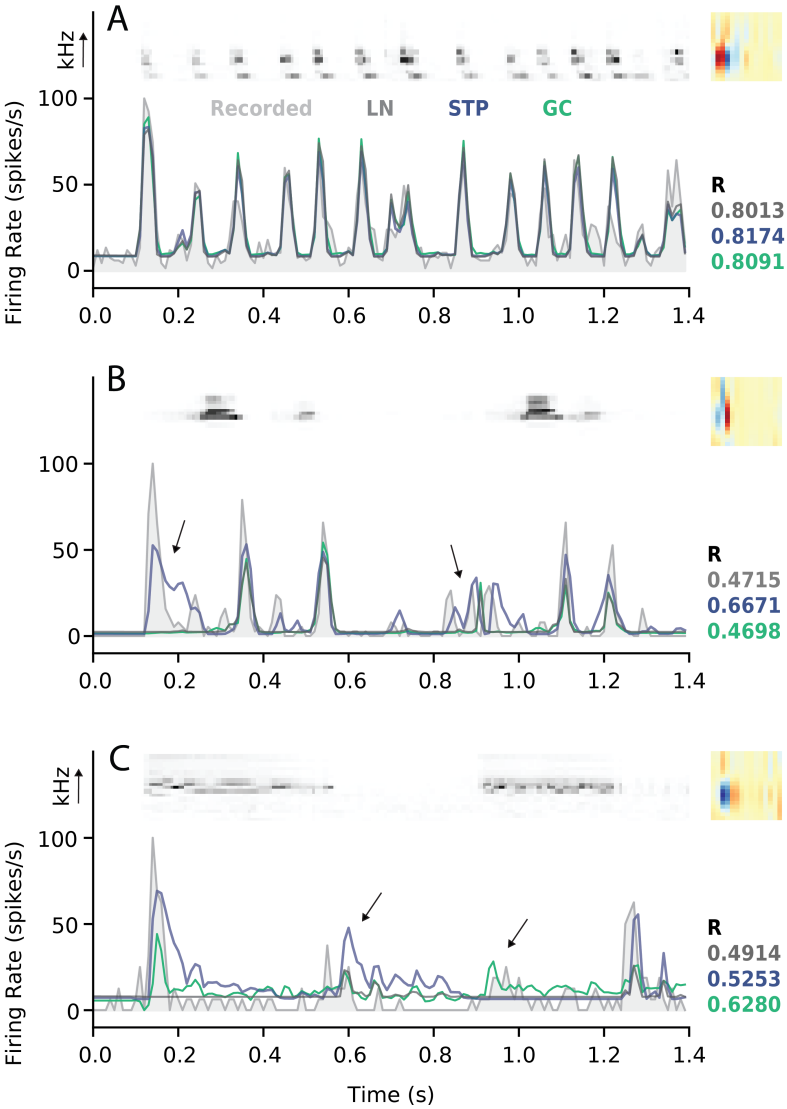

Figure 3: Example model fits and predictions. (A) Results from a neuron for which the STP and GC models predictions were not significantly better than the LN model prediction. Top left subpanel shows the spectrogram from one natural sound in the validation set. Top right panel shows the spectro-temporal filter from the LN model fit (right). Bottom panel shows the actual response (light gray, filled) overlaid with predictions by the LN (dark gray), STP (blue), and GC (green) models. Values at the bottom right of each panel are the prediction correlations for each model. The actual response was smoothed using a 30ms box filter for easier visualization. (B) Data from a neuron for which the STP model performed significantly better than the LN and GC models, plotted as in A. Arrows indicate times for which the STP model successfully reproduced an increase in firing rate while the other models did not. (C) Data from a neuron for which the GC model performed significantly better than the LN and STP models. Arrows indicate a time for which the STP model incorrectly predicted an increase in firing rate (left) and a time for which the GC model successfully reproduced an increase in firing rate while the other models did not (right).

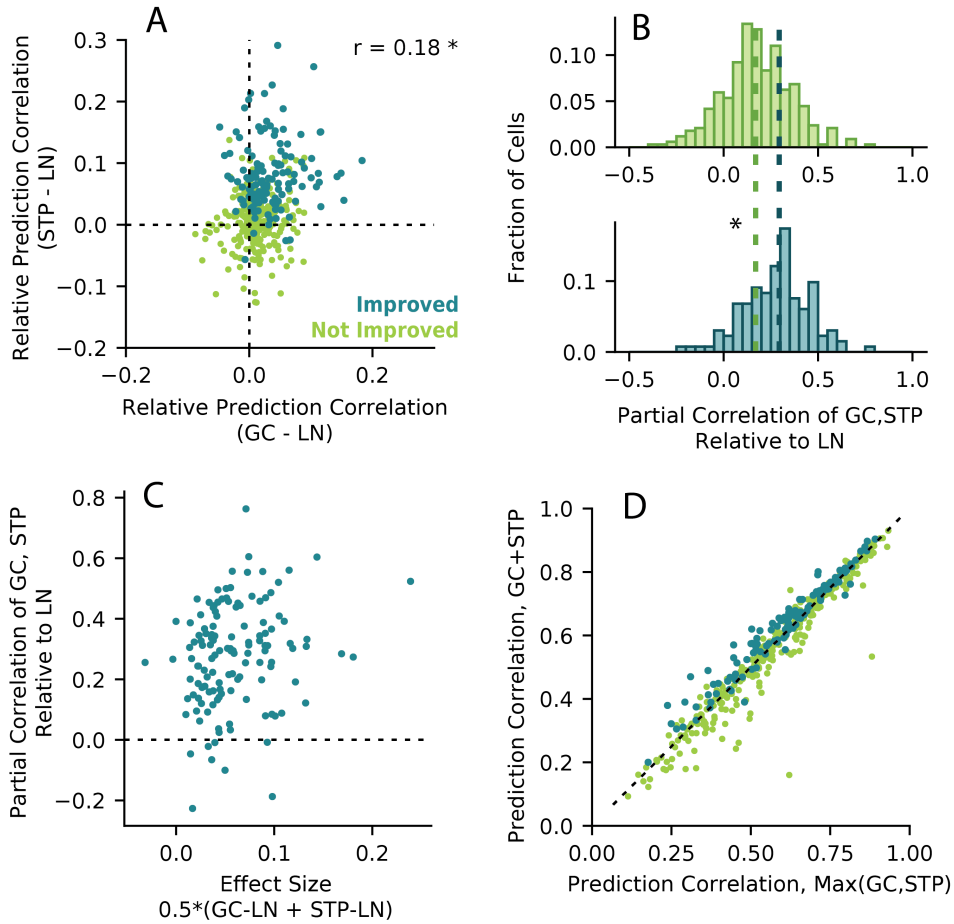

Figure 4: Equivalence of model predictions. (A) Change in prediction correlation for the GC (horizontal axis) and STP (vertical axis) models relative to the LN model for each neuron ($r = 0.18, p = 6.42 \times 10^{-5}$). Blue points indicate neurons with a significant improvement for the GC+STP model over the LN model ($p < 0.05$, permutation test); green points indicate neurons that were not improved. (B) Equivalence scores, measured as the partial correlation between STP and GC model predictions relative to the LN model prediction. Median equivalence for improved cells (0.29) was significantly greater than for non-improved cells (0.17) (Mann-Whitney U test, $p = 1.72 \times 10^{-8}$). (C) Scatter plot compares equivalence (vertical axis) versus effect size (horizontal axis), *i.e.*, the average change in prediction correlation for the STP and GC models relative to the LN model. No relationship between equivalence and effect size was observed. (D) Prediction correlations for the combined GC+STP model (vertical axis) and the maximum of the GC and STP models (horizontal axis). The median prediction correlation for the combined model was significantly higher than the greater of the individual models (Wilcoxon signed rank test, $p = 0.0444$)

and GC model predictions that negligibly differed from the LN model prediction (*e.g.*, Fig. 3a). Such predictions could be called trivially equivalent in that they offer no additional explanatory power, and calling two models equivalent based on such comparisons would not be very useful. We therefore defined a measure of effect size for each cell as the mean change in prediction correlation for the STP and GC models relative to the LN model. Neurons for which the more complex models did not improve prediction accuracy compared to the LN model might have a high equivalence score, but would also have a small effect size. If the STP and GC models were non trivially equivalent, we would expect most cells with significant improvements over the LN model to only have a large effect size if they also had a high equivalence score (but not necessarily the reverse). However, we did not discern any such pattern in the data (Fig. 4c). Instead, equivalence and effect size were mostly unrelated.

Following this evidence for limited equivalence, we wanted to know whether the combined model's greater predictive power was merely the result of its ability to simply account for either STP or GC, without any benefit from their combination in individual neurons. If this were the case, we would expect that for any given cell, the prediction correlation of the combined model should be no greater than the larger of the prediction correlations for the STP and GC models (Fig. 4d). Instead, we found that the median prediction correlation was significantly higher for the combined model, although the difference was small (Wilcoxon signed rank test, two-sided, $p = 0.0444$, median difference 0.001). We also observed that the responses of improved cells were almost exclusively predicted better by the combined model than by either the STP or GC model individually. This indicated to us that a small number of neurons exhibit dynamics explained better by a combination of the STP and GC nonlinearities. However, the majority of the improved neurons appeared to utilize only one mechanism or the other.

Model fit parameters are consistent with previous measurements.

Since both the STP model and the GC model used in this study were designed to replicate previous work (Lopez Espejo, Schwartz, and David 2019; Rabinowitz, Willmore, Schnupp, et al. 2012), we wanted to verify that the models behaved in a manner consistent with previous observations. This was of particular concern for the GC model since it had not previously been fit using a natural sound dataset. To do this, we analyzed the distributions of their fitted parameter values for improved versus non-improved cells.

For the short-term plasticity model (Fig. 5), we found that the median value of both the time constant (τ) and fraction gain change (u) parameters was significantly increased for improved versus non-improved cells (Mann-Whitney U tests, two-sided, $p = 3.53 \times 10^{-6}$ and $p = 2.92 \times 10^{-13}$, respectively). However, the difference was more pronounced for the u parameter. Additionally, nearly all cells had positive values for the u parameter, indicating predominant depression rather than facilitation, which agreed with published results (Lopez Espejo, Schwartz, and David 2019).

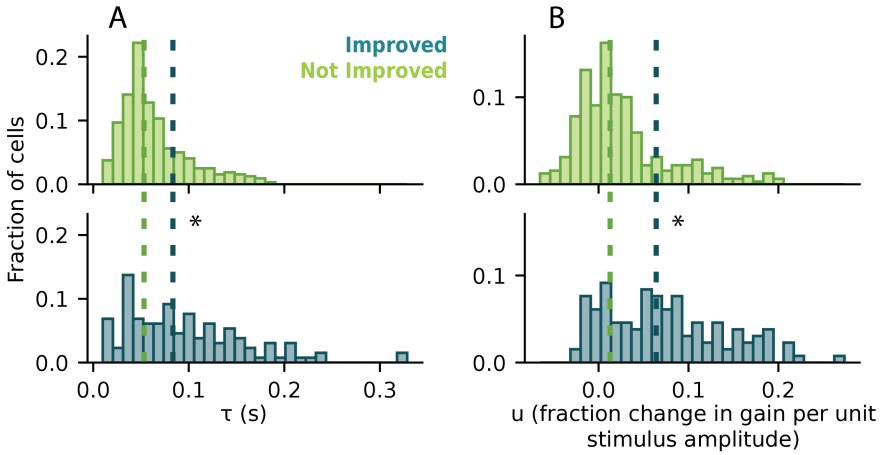

Figure 5: Parameter fit values for the short-term plasticity model. (A) Distribution of τ, representing the time constant for the recovery of synaptic vesicles. Top panel shows data for non-improved neurons and bottom panel for improved neurons. Median values for non-improved (0.0533) and improved (0.0833) neurons were significantly different ($p = 3.53 \times 10^{-6}$, two-sided Mann-Whitney U test, * $p < 0.05$), indicating a longer time constant for the improved cells. (B) Distribution of u values, representing release probability (*i.e.*, the fraction change in gain per unit of stimulus amplitude). Medians for non-improved (0.0128) and improved (0.0641) neurons were significantly different from one another ($p = 2.92 \times 10^{-13}$), showing higher release probability for improved neurons.

For the contrast-dependent gain control model (Fig. 6), we found that the slope (k) parameter of the static nonlinearity decreased for high contrast sounds (Mann-Whitney U tests, two-sided; $p = 1.22 \times 10^{-4}$). This change was consistent with models fit using random contrast dynamic random chord (RC-DRC) stimuli (Rabinowitz, Willmore, Schnupp, et al. 2012). Meanwhile, the saturation (a) decreased and baseline (b) increased with contrast ($p = 8.01 \times 10^{-6}$ and $p = 5.90 \times 10^{-7}$, respectively). However, we observed no significant effect for the input offset (s) parameter, which did also change for RC-DRC stimuli. As in the previous study, the net result of increasing contrast was to decrease the gain of neural responses, so the overall effects of changing contrast on response gain were consistent.

Simulations suggest distinct functional domains for the short-term plasticity and contrast-dependent gain control models.

The weak equivalence of the GC and STP models reported above supports a conclusion that these models account for functionally distinct response properties in A1. However, the low equivalence could result from disparate degrees of optimization between models across cells. If this were the case, then models fit to simulated, noise-free data should have a higher degree of equivalence. To address this concern, we fit the LN, STP and GC models to three simulated neural responses (Fig. 7). Each simulation was generated by using a fitted model to predict responses to a set of natural stimuli and treating the model's prediction as ground-truth for subsequent fitting. The first simulated response was generated using the

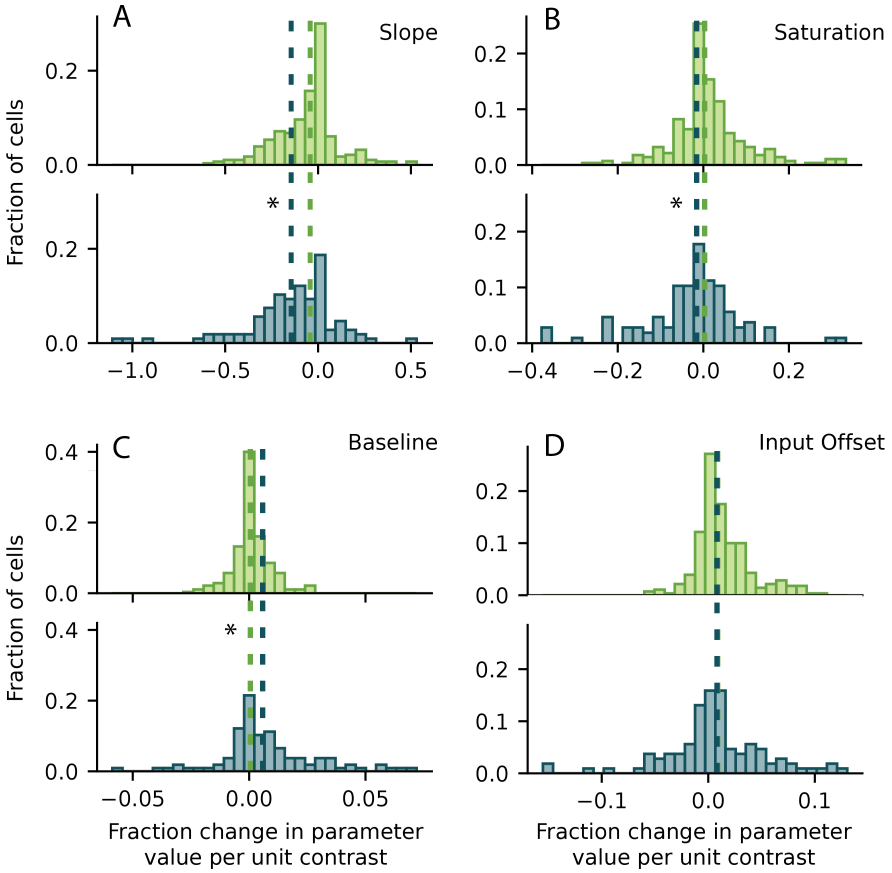

Figure 6: Parameter fit values for the contrast-dependent gain control model. (A) Histogram of effect of contrast on k, representing the slope of the output nonlinearity. Top panel shows data for non-improved neurons (green) and bottom panel for improved neurons (blue). The negative median value for improved (-0.14) cells indicates a decrease in slope during high-contrast conditions. This median was significantly smaller than for non-improved neurons (-0.042) ($p = 1.22 \times 10^{-4}$, two-sided Mann-Whitney U test). Asterisk (*) indicates $p < 0.05$. (B) Histogram of contrast effect on a (saturation level), plotted as in A. The median parameter values for non-improved (0.0031) and improved (-0.0156) neurons were significantly different ($p = 8.01 \times 10^{-6}$), indicating a decreased response amplitude in high contrast conditions. (C) Distribution of contrast effect on b (baseline of the output nonlinearity). Medians for non-improved (0.0005) and improved (0.0058) neurons were significantly different from one another ($p = 5.90 \times 10^{-7}$), indicating an increase in baseline for high contrast. (D) Distribution of contrast effect on s (input offset). There was no significant difference between medians for non-improved (0.0088) and improved (0.0082) neurons ($p = 0.85$).

LN model fit to a cell that showed no improvements for the STP or GC models (Fig. 3a). The other simulations were generated using the STP and GC model fits from Fig. 3b and c, respectively. These latter two simulations were chosen to represent neurons for which either the STP or GC model performed better than the LN model.

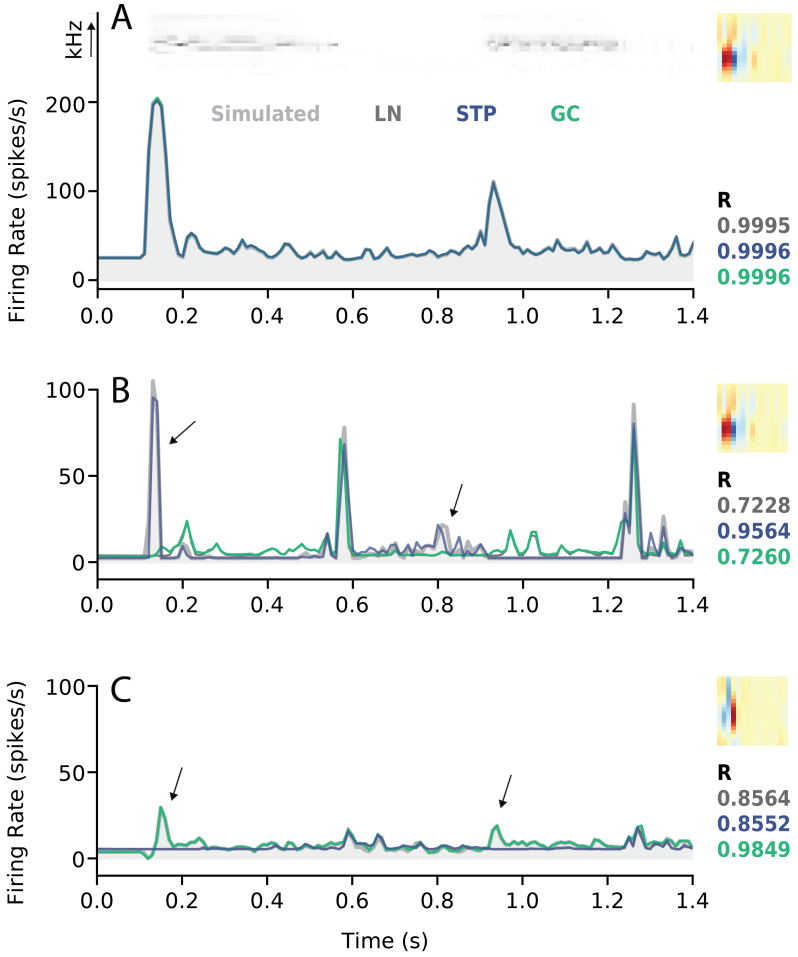

Figure 7: Model performance for simulated data. Each panel shows the predicted response by each model to one stimulus (spectrogram at top). In each case, a different type of neuron (LN, STP or GC) was used to simulate responses, using parameter values fit from real examples (Figure 3). Then each of the three models was fit to the simulated data. The linear filter fit using the LN model is shown at the top right of each panel. The prediction correlations for each model (across all stimuli) are shown at the bottom right (LN: dark gray, STP: blue, GC: green). (A) Simulation based on the fitted LN model from Fig. 3a. (B) Simulation based on the fitted STP model from Fig. 3b. (C) Simulation based on the fitted GC model from Fig. 3c.

As expected, all three models were able to reproduce the LN simulation nearly perfectly (Pearson's R = 0.9995, 0.9996, and 0.9996 for the LN, STP, and GC models, respectively). However, when fit to the STP simulation, the GC model was no better than the LN model, but the STP model did perform better (R = 0.7228, 0.9564, and 0.7260). Conversely, when fit to the GC simulation, the STP model performed no better than the LN model while the

GC model did (R = 0.8564, 0.8552, 0.9849). This pattern confirmed that that the STP and GC models did indeed capture distinct neuronal dynamics.

Greater relative contribution of contrast gain to encoding of noisy natural sounds.

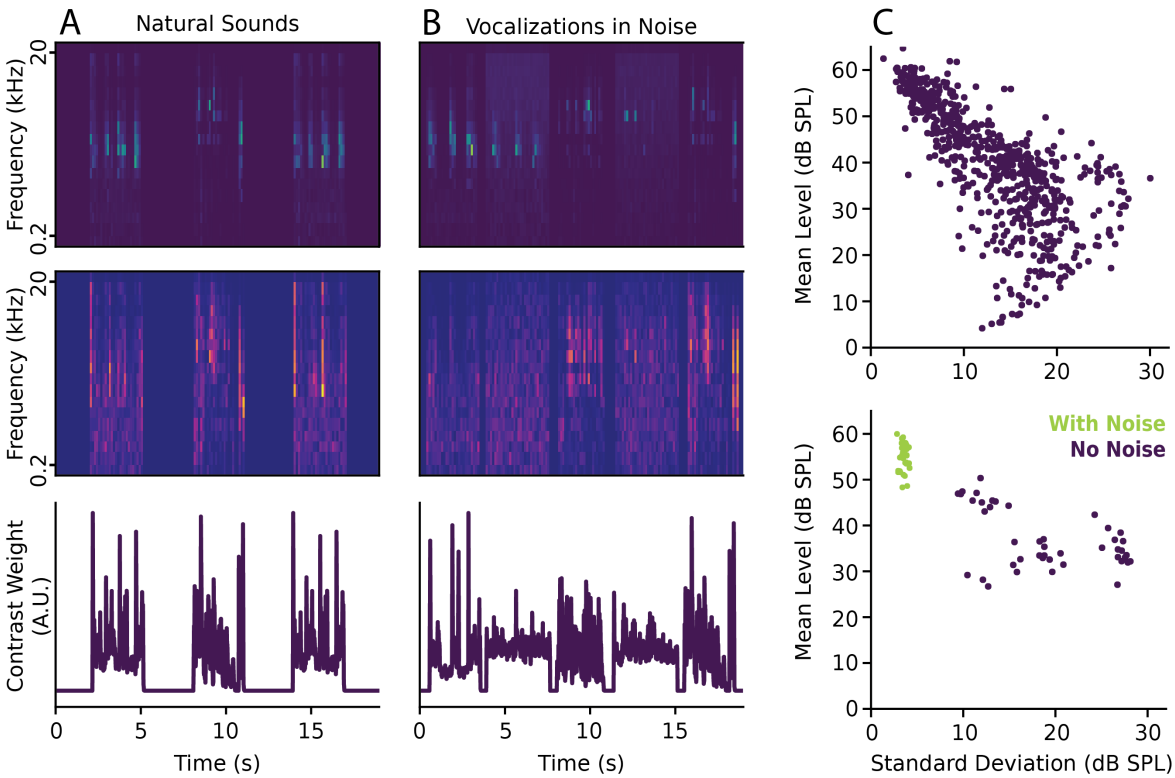

Figure 8: Distribution of contrast level for clean and noisy stimuli. (A) From top to bottom: stimulus spectrogram, contrast, and frequency-summed contrast for three natural sound samples. (B) Same as in A, but for five vocalizations with broadband noise added to the second and fourth samples. (C) Standard deviation and mean level (dB SPL) for all natural sounds (top) and clean/noisy vocalizations (bottom) from each dataset. For the natural sound set, the distribution smoothly varies across contrast levels. For the noisy stimuli, there is a clear grouping of noisy (low-) and clean (high-contrast) stimuli.

In addition to natural sounds, we also compared performance of the STP and GC models on data collected with clean and noisy vocalizations. Previous studies using stimulus reconstruction methods argued that both short-term plasticity and contrast-dependent gain control are necessary for robust encoding of noisy natural signals (Mesgarani et al. 2014; Rabinowitz, Willmore, King, et al. 2013). The inclusion of additive noise is likely to reduce contrast by increasing the mean stimulus energy without changing variance. We compared the mean and standard deviation of each sample for the two data sets to see if there were any systematic differences in contrast. We found that the natural sounds dataset smoothly

spanned the range of observed means and standard deviations. Meanwhile, the dataset containing clean and noisy vocalizations naturally formed two distinct categories categories: high-contrast (clean) and low-contrast (noisy) (Fig. 8).

We compared model performance and equivalence, using the same approach as for the natural sound data above (Fig. 9). For the noisy vocalization data, we again found that both the STP and GC models performed significantly better than the LN model and that the combined model performed even better (Wilcoxon signed-rank test, two-sided, $p = 1.7 \times 10^{-8}$, $p = 0.0058$, $p = 4.3 \times 10^{-5}$, and $p = 3.7 \times 10^{-9}$, respectively). However, unlike for the natural sound data, performance of the STP and GC models themselves was not significantly different (Wilcoxon signed-rank test, two-sided, $p = 0.11$). This indicated a relative increase in the performance of the GC model when applied to noisy vocalizations, likely owing to the large fluctuations between between high- and low-contrast stimuli.

The results of the equivalence analyses were also similar between data sets. For the noisy vocalizations, the change in prediction correlation for the STP and GC models was only weakly correlated ($r = 0.18$, $p = 3.5 \times 10^{-}2$). Also, partial correlation between STP and GC predictions was modest for improved cells (median 0.30), but significantly higher than for non-improved cells (median 0.18) (Mann-Whitney U test, two-sided, $p = 0.047$). However, the small number of significantly improved cells in this dataset ($n = 12/141$) made drawing definitive conclusions difficult.

Discussion

We find that encoding models incorporating either gain control (GC) or short-term synaptic plasticity (STP) explain complementary nonlinear components of neural responses in primary auditory cortex (A1). Although we observe some overlap in explanatory power between models, their equivalence is relatively weak (Fig. 4). Instead, a novel model which incorporates both STP and GC mechanisms shows improved performance over either separate model (Fig. 2). It is well-established that the LN model fails to account for important contextual effects. This work supports the idea that both forward adaptation, mediated by a mechanism such as STP, and feedback inhibition, as might be mediated by GC, both play a role in these contextual processes.

Comparison with previous models

There are some important differences between the GC model implementation used in this study and the original model implemented in Rabinowitz, Willmore, Schnupp, et al. 2012. First, whereas Rabinowitz et al. imposed the contrast profiles of their stimuli by design, natural stimuli contain dynamic fluctuations in contrast that must be calculated heuristically. As a result, we were forced to make decisions about parameters governing the contrast

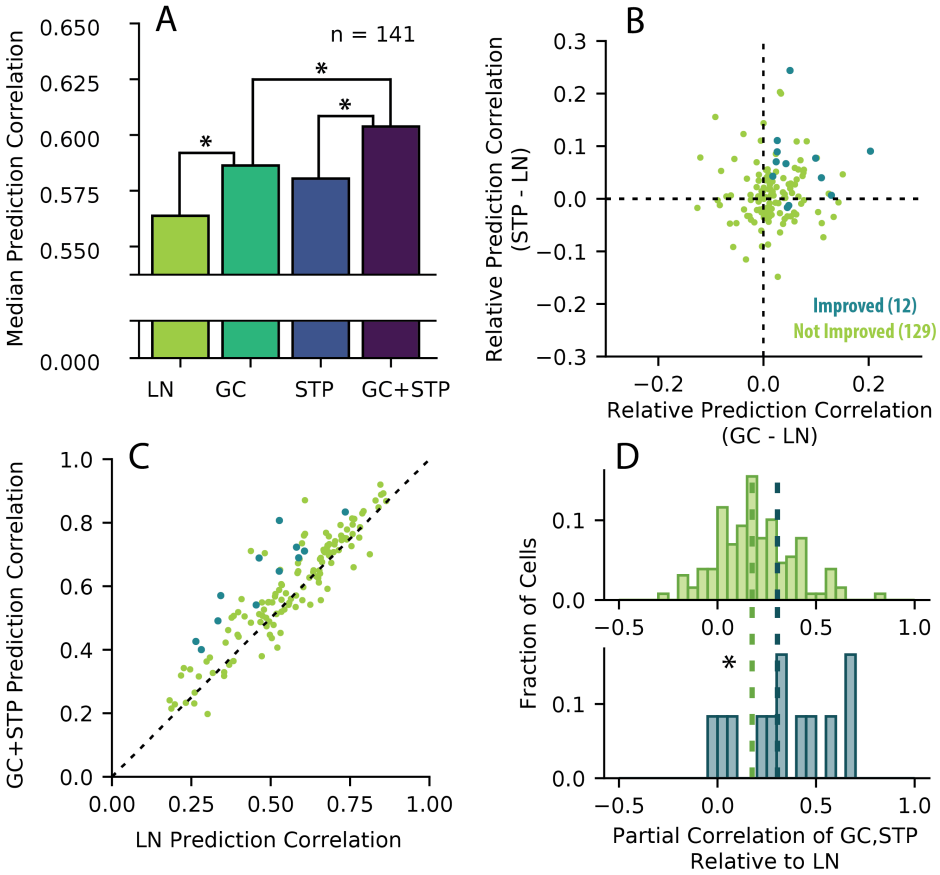

Figure 9: Comparison of model performance for data including clean and noisy vocalizations. (A) Median prediction correlation ($n = 141$) for each model. Differences were significant between LN and GC models ($p = 1.70 \times 10^{-8}$), GC and GC+STP models ($p = 4.27 \times 10^{-5}$), and STP and GC+STP models ($p = 3.69 \times 10^{-9}$, two-sided Wilcoxon signed-rank test. However, unlike the natural sound data (Fig. 2), performance was not significantly different between the GC and STP models. (B) Change in prediction correlation for the GC (horizontal axis) and STP (vertical axis) models relative to the LN model for each neuron ($r = 0.18, p = 3.45 \times 10^{-2}$). (C) Prediction correlations for the LN model (horizontal axis) compared to the combined model (vertical axis) for each neuron, grouped by whether the combined model showed a significant improvement (blue, $n = 12$, $p < 0.05$, permutation test) or not (green, $n = 129$). (D) Histogram of equivalence for non-improved (top) and improved neurons (bottom). Median equivalence for improved cells (0.30) was significantly greater than for non-improved cells (0.18, Mann-Whitney U test, $p = 0.047$).

metric: the spectral and temporal extent of the window used to calculate contrast and the temporal offset needed to emulate the dynamics with which contrast effects are fed back to the response. Experimentally, we found that a 70ms, spectrally narrowband convolution window and 20ms temporal offset worked best across our dataset on average. However, model performance may be further improved if these parameters are optimized on a cell-by-cell basis. For the STP model, optimal plasticity parameters did vary between cells. A variability in contrast parameters may reflect similar biologically relevant differences in the ways that cells adapt to contrast. A second important difference from the original GC model is that we were not able to differentiate high contrast sounds with high standard deviation from those with low mean level—both cases result in a larger coefficient of variation. In the original study, Rabinowitz et al. were able to fix mean level across stimuli in order to avoid this potential confound (Rabinowitz, Willmore, Schnupp, et al. 2012). Despite these differences, however, our results broadly replicated the original findings.

Mechanisms mediating effects of sensory context on auditory cortical responses

Because the current study focuses on functional models of cortical activity without direct manipulation of neural circuits, we cannot be certain how our modeling results translate to different mechanisms in the brain. It is possible that the STP and GC models are both describing synaptic adaptation, but that network- or cell-level features make one or the other model more suited to predicting the responses of one group of cells over another. Direct manipulations of synaptic plasticity and/or inhibitory feedback mechanisms can provide explicit insight into the mechanisms underlying these functions. However, regardless of the underlying physiology, it is clear that both models are needed to fully characterize the responses of a diverse population of auditory neurons.

References

Aertsen, A. M. and P. I. Johannesma (1981). "The spectro-temporal receptive field: a functional characteristic of auditory neurons." In: *Biological Cybernetics* 42, pp. 133–143.

Calabrese, Ana et al. (2011). "A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds." In: *PLOS ONE*. DOI: 10.1371/journal.pone.0016104.

Carandini, Matteo, David Heeger, and Walter Senn (2002). "A Synaptic Explanation of Suppression in Visual Cortex." In: *The Journal of Neuroscience* 22 (22), pp. 10053–10065. DOI: 10.1523/JNEUROSCI.22-22-10053.2002.

Cooke, James et al. (2018). "Contrast gain control in mouse auditory cortex." In: *Journal of Neurophysiology* 120, pp. 1872–1884. DOI: 10.1152/jn.00847.2017.

David, Stephen V. (2018). "Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding." In: *Hearing Research* 360, pp. 107–123. DOI: 10.1016/j.heares.2017.12.021.

David, Stephen V., Nima Mesgarani, et al. (2009). "Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli." In: *The Journal of Neuroscience* 29 (11), pp. 3374–3386. DOI: 10.1523/JNEUROSCI.5249-08.2009.

David, Stephen V. and Shihab A. Shamma (2013). "Integration over multiple timescales in primary auditory cortex." In: *The Journal of Neuroscience* 33 (49), pp. 19154–19166. DOI: 10.1523/JNEUROSCI.2270-13.2013.

Englitz, B. et al. (2013). "MANTA - an open-source, high density electrophysiology recording suite for MATLAB." In: *Frontiers in neural circuits* 7, p. 69. DOI: 10.3389/fncir.2013.00069.

Hiroki, Asari and Anthony Zador (2009). "Long-lasting context dependence constrains neural encoding models in rodent auditory cortex." In: *Journal of Neurophysiology* 102, pp. 2638–2656. DOI: 10.1152/jn.00577.2009.

Hsu, A., A. Borst, and F. Theunissen (2004). "Quantifying the variability in neural responses and its application for the validation of model predictions." In: *Network* 15, pp. 91–109. DOI: 10.1088/0954-898X/15/2/002.

Katsiamis, A., E. Drakakis, and R. Lyon (2007). "Practical gammatone-like filters for auditory processing." In: *EURASIP Journal on Audio, Speech and Music Processing* 2007.137 (1).

Kowalski, N., D. Depireux, and S. Shamma (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra." In: *Journal of Neurophysiology* 76 (5), pp. 3503–3523.

Lopez Espejo, Mateo, Zachary Schwartz, and S. V. David (2019). "Spectral tuning of adaptation supports coding of sensory context in auditory cortex." In: *PLOS Computational Biology* 15.10. DOI: 10.1371/journal.pcbi.1007430.

McDermott, Josh, Michael Schemitsch, and Eero Simoncelli (2013). "Summary statistics in auditory perception." In: *Nature* 16 (4), pp. 493–498. DOI: 10.1038/nn.3347.

Mesgarani, Nima et al. (2014). "Mechanisms of noise robust representation of speech in primary auditory cortex." In: *PNAS* 111 (18), pp. 6792–6797. DOI: 10.1073/pnas.1318017111.

Rabinowitz, N., B. Willmore, A. King, et al. (2013). "Constructing noise-invariant representations of sound in the auditory pathway." In: *PLOS Biology* 11 (11), e1001710. DOI: 10.1371/journal.pbio.1001710.

Rabinowitz, N., B. Willmore, J. Schnupp, et al. (2011). "Contrast gain control in auditory cortex." In: *Neuron* 70 (6), pp. 1178–1191. DOI: 10.1016/j.neuron.2011.04.030.

– (2012). "Spectrotemporal contrast kernels for neurons in primary auditory cortex." In: *The Journal of Neuroscience* 32 (33), pp. 11271–11284. DOI: 10.1523/JNEUROSCI.1715-12.2012.

Rahman, Monzilur et al. (2019). "A dynamic network model of temporal receptive fields in primary auditory cortex." In: *PLOS Computational Biology*. DOI: 10.1371/journal.pcbi.1006618.

Simoncelli, Eero et al. (2003). "Characterization of Neural Responses with Stochastic Stimuli." In:

Thorson, I., J. Liénard, and S. David (2015). "The essential complexity of auditory receptive fields." In: *PLOS Computational Biology* 11 (12), e1004628. DOI: 10.1371/journal.pcbi.1004628.

Tsodyks, M., K. Pawelzik, and H. Markram (1998). "Neural networks with dynamic synapses." In: *Neural Computation* 10 (4), pp. 821–835. DOI: 10.1162/089976698300017502.

Wehr, Michael and Anthony Zador (2005). "Synaptic mechanisms of forward suppression in rat auditory cortex." In: *Neuron* 47, pp. 437–445. DOI: 10.1016/j.neuron.2005.06.009.