

**Title**

A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds.

**Authors**

Ana Calabrese<sup>1,2</sup>, Joseph W. Schumacher<sup>1</sup>, David M. Schneider<sup>1</sup>, Liam Paninski<sup>1,2,3</sup> and Sarah M.N. Woolley<sup>1,4,\*</sup>

**Affiliations**

<sup>1</sup> Doctoral Program in Neurobiology and Behavior, Columbia University, New York, New York

<sup>2</sup> Center for Theoretical Neuroscience, Columbia University, New York, New York

<sup>3</sup> Department of Statistics, Columbia University, New York, New York

<sup>4</sup> Department of Psychology, Columbia University, New York, New York

**Running title**

GLM with natural sounds

**\*Correspondence:**

Sarah M. N. Woolley  
Columbia University  
1406 Schermerhorn Hall  
1190 Amsterdam Ave.  
New York, NY 10027  
phone: 212-854-5448  
fax: 212-854-3609  
email: sw2277@columbia.edu

## ABSTRACT

In the auditory system, the stimulus-response properties of single neurons are often described in terms of the spectrotemporal receptive field (STRF), a linear kernel relating the spectrogram of the sound stimulus to the instantaneous firing rate of the neuron. Several algorithms have been used to estimate STRFs from responses to natural stimuli; these algorithms differ in their functional models, cost functions, and regularization methods. Here, we characterize the stimulus-response function of auditory neurons using a generalized linear model (GLM). In this model, each cell's input is described by: 1) a stimulus filter (STRF); and 2) a post-spike filter, which captures dependencies on the neuron's spiking history. The output of the model is given by a series of spike trains rather than instantaneous firing rate, allowing the prediction of spike train responses to novel stimuli. We fit the model by maximum penalized likelihood to the spiking activity of zebra finch auditory midbrain neurons in response to conspecific vocalizations (songs) and modulation limited (ml) noise. We compare this model to normalized reverse correlation (NRC), the traditional method for STRF estimation, in terms of predictive power and the basic tuning properties of the estimated STRFs. We find that a GLM with a sparse prior predicts novel responses to both stimulus classes significantly better than NRC. Importantly, we find that STRFs from the two models derived from the same responses can differ substantially and that GLM STRFs are more consistent between stimulus classes than NRC STRFs. These results suggest that a GLM with a sparse prior provides a more accurate characterization of spectrotemporal tuning than does the NRC method when responses to complex sounds are studied in these neurons.

## KEYWORDS

STRF; auditory midbrain; GLM; natural sounds; zebra finch; sensory coding.

## INTRODUCTION

Characterizing neural responses to natural stimuli remains one of the ultimate goals of sensory neuroscience. However, considerable technical difficulties exist for correctly estimating neural receptive fields (RFs) from natural stimuli. Two major difficulties are the interactions between higher-order statistics of the stimuli and inherent nonlinearities of neural responses (Sharpee et al. 2008, Christianson et al. 2008) and the challenge of estimating receptive fields in high dimensional spaces with limited data (Theunissen et al. 2001, David et al. 2007).

Neural responses are commonly characterized by a linear-nonlinear (LN) model (Chichilnisky 2001, Paninski 2003, Simoncelli et al. 2004), in which the output of a linear filter or receptive field (RF) applied to the stimulus is then transformed by a static nonlinearity to determine the instantaneous firing rate of the neuron. Reverse correlation (RC), the most widely used estimation method, computes the RF of a neuron by multiplying the spike-triggered average (STA) of the stimulus by the inverse of the stimulus covariance matrix. It is well understood that, for an LN neuron, RC is guaranteed to produce an unbiased estimate of a neuron's true underlying filter only if the distribution of the stimuli used for estimation is elliptically symmetric (Paninski 2003). Deviations from either the LN framework (e.g., the existence of more than one linear filter (multiple-filter LN), or extra terms that take into account spiking history) or the elliptical symmetry condition (e.g., naturalistic stimuli which contain higher order correlations) can introduce biases in the estimate of the RF.

The highly correlated structure of natural stimuli presents additional numerical problems for RF estimation. Because natural stimuli contain strong autocorrelations, the majority of the power in the stimulus tends to be concentrated in a small number of dimensions. Multiplication by the inverse of the stimulus covariance matrix causes noise in the resulting RF to be strongly amplified along the stimulus dimensions with low variance. Thus, some form of regularization is applied to RC to prevent overfitting to noise (Theunissen et al. 2001, Sahani and Linden 2003, Smyth et al. 2003, Machens et al. 2003).

In the auditory system, the stimulus-response properties of single neurons are often described in terms of the spectrotemporal receptive field (STRF), a linear kernel relating the spectrogram of the sound stimulus to the instantaneous firing rate of the neuron. Traditionally, STRFs have been estimated using normalized-reverse correlation (NRC), a method that uses an approximation to the stimulus covariance matrix to obtain regularized estimates. Regularization introduces a prior that imposes constraints in the STRF estimate and, under noisy conditions, the specific regularization used by the model can introduce biases in the estimates (David et al. 2007). Methods other than the NRC have recently been proposed to characterize the tuning properties of auditory neurons from responses to natural stimuli, each of which reduces the impact of

stimulus-correlation biases on the estimated STRFs (David et al. 2007, Sahani and Linden 2003, Sharpee et al. 2004). These algorithms differ in their functional models, cost functions, and regularization methods. Here, we propose an approach for characterizing the stimulus-response function of auditory neurons based on a generalized linear model (GLM). This method is advantageous because it requires relatively light computational resources and provides easily interpretable results (Paninski 2004, Truccolo et al. 2005, Brillinger 1988). For example, it has been successfully used to accurately predict spiking responses (Pillow et al. 2005) of single retinal ganglion cells as well as detailed spatiotemporal correlations in the responses of a complete population of macaque retinal ganglion cells (Pillow et al. 2008). As opposed to most STRF estimation methods, our method takes into account spiking history. Further, the output of the model is a series of spike trains rather than average time-varying firing rate, allowing comparison of the actual and predicted spike train responses.

In this study, we compare a GLM with a sparse prior and NRC in terms of their ability to predict responses to novel stimuli and the tuning properties of the STRFs they produce. We fit both models to responses of single auditory neurons in the midbrain of zebra finches probed with two types of stimuli: zebra finch songs and modulation-limited noise. We find that the GLM predicts responses to both stimulus classes significantly better than NRC, and that GLM and NRC STRFs derived from the same data can differ profoundly. Finally, the GLM method reduces differences in tuning between stimulus classes.

## MATERIALS AND METHODS

### 2.1 *Electrophysiology*

All procedures were in accordance with the NIH and Columbia University Animal Care and Use Policy. Electrophysiological recordings were made from single neurons in the auditory midbrain, mesencephalic lateralis dorsalis (MLd), of adult male zebra finches as described in (Schneider and Woolley 2010). Briefly, an initial preparatory surgery was performed 48 hrs before the first neural recording session. Birds were deeply anesthetized with 0.03 ml Equithesin and placed in a custom stereotaxic holder. For recordings made from anesthetized birds, only the first layer of skull was removed during the initial surgery. For recordings made from awake birds, full craniotomies were made. A grounding wire was cemented in place with its end just beneath the skull, approximately 5 to 10 mm lateral to the junction of the midsagittal sinus. A head post was cemented to the skull of the animal and points were marked for electrode penetrations. Anesthetized recording sessions were preceded by administering three doses of 0.03 ml of 20% urethane over a period of one hour. Recordings were made using glass pipettes containing 1M NaCl, with impedances ranging from 5 to 20 MOhms. The duration of the recording sessions

ranged from 4 to 15 hours. Awake recording sessions were no longer than 6 hours. For a single animal, awake recordings were performed over a period of approximately two weeks and anesthetized recordings were performed in a single session. After final recording sessions, the birds were euthanized and brains were preserved for histological reconstruction of electrode locations.

## 2.2 *Stimuli*

Two sound ensembles were used: a conspecific song ensemble and a modulation-limited (ml) noise ensemble. The conspecific song ensemble consisted of 20 songs (~ 2 sec in duration each) from different adult male zebra finches. Each song was band-pass frequency filtered between 250 Hz and 8 kHz. The ml noise ensemble consisted of 10 samples of 2 sec of ml noise. MI noise is a behaviorally meaningless sound similar to white noise that was designed to match song in frequency range, maximum spectral and temporal modulations and power (Woolley et al. 2005). Stimuli were presented at a mean intensity of 72 dB SPL. Ten spike train response trials were obtained for each of the 20 songs and 10 noise samples. Trials for different stimuli were interleaved in random order. The inter-trial interval was determined at random from a uniform distribution between 1.2 and 1.6 seconds.

## 2.3 *Data preprocessing*

The same preprocessing was applied to the data before fitting both NRC and GLM. Spectrograms were generated from the stimulus sound pressure waveforms using a bank of band-pass filters with center frequencies ranging from 250 to 8000 Hz, which covers the audible frequency range for zebra finches (Zevin et al. 2004). The center frequencies were spaced linearly and had a bandwidth of 125 Hz. It has been shown that the predictive abilities of STRFs can be improved by applying a compressive nonlinearity to the stimulus spectrogram (Gill et al. 2006). We therefore applied a logarithm to the stimulus spectrogram prior to fitting the models, which mimics peripheral auditory processing.

For the NRC method, both stimulus spectrograms and spike trains were binned at 1 ms resolution (the temporal resolution required by STRFPak, the publicly available Matlab toolbox for STRF estimation we used in this study; see Section 2.5). For the GLM method, both signals were further down sampled by a factor of 3. Using time bins larger than 1 ms is common in the GLM setting (Eldawlatly et al. 2009). Expanding the bin size can avoid nonconvergence problems related to the refractory periods of neurons (Zhao and Lyendar 2010), and effectively reduces the computational load. In order to test that the different bin sizes in the estimation of NRC and GLM STRFs was not introducing a bias in predictive power or STRF shape, we re-computed the STRFs of a subset of our population of cells (10 cells) using the GLM with 1 ms time bins. We

found no significant differences in STRF shape or predictive power of GLM STRFs computed at 1 ms or 3 ms resolution (the average same-class prediction correlation for a novel song was 0.507 and 0.51, respectively).

#### 2.4 Generalized linear model for spike trains

We describe the encoding of a stimulus spectrogram (a transformation of the sound pressure waveform into a time-varying function of energy at each frequency band),  $\bar{X}$ , in the spike trains of an auditory neuron with a generalized linear model (GLM, Fig. 2A), a generalization of the well known linear-nonlinear-Poisson (LNP) cascade model (Paninski 2004). In this model, a cell's response is described by: 1) a stimulus filter, or STRF ( $\bar{k}$ ); 2) a post-spike filter ( $\bar{h}$ ), which captures dependencies on the cell's spiking history (e.g. refractoriness); and 3) a constant offset  $b$  which sets the threshold and baseline firing rate of the model. For each neuron, a static nonlinear function is then applied to the summed filter responses to obtain an instantaneous spike rate (Truccolo et al. 2005, Brillinger 1988, Paninski et al. 2007). Although these types of models are strictly phenomenological, their components can be broadly compared to biophysical mechanisms. The stimulus filter approximates the spectrotemporal integration of the sound stimulus in an auditory neuron. The post-spike filter mimics voltage-activated currents following a spike. And the output nonlinearity implements a soft threshold converting membrane potential to instantaneous spike probability.

##### 2.4.1 Model fitting/parameter estimation

We fit the model to extracellular single unit recordings from 169 auditory midbrain neurons. To calculate the model parameters, stimuli (log spectrograms) were computed and responses (spike trains) were binned at a 3 ms resolution. The time bin was small enough that more than one spike was almost never observed in any bin. The spectral domain of the stimulus was divided into 20 equally spaced frequency bins, which spanned frequencies between 250 and 8000 Hz. The model parameters were fitted by maximum penalized likelihood. The parameters for each cell consisted of a stimulus filter or STRF  $\bar{k}$ , a constant offset  $b$ , and a post-spike filter  $\bar{h}$ . The STRF was a 400-dimensional vector (20 spectral  $\times$  20 time bins, including frequencies between 250 and 8000 Hz and latencies between 0 and 60 ms, respectively), the post-spike filter was a 5-dimensional vector (5 time bins spanning the 15 ms following each spike) and the offset consisted of a scalar value, for a total of 406 unknown parameters. Increasing the binning resolution would change the number of fit parameters and could, in theory, improve the performance of the model. However, the resolution used in this study is sufficient for demonstrating the performance of GLM compared to NRC under these experimental conditions.

The conditional spike rate in the model is given by

$$\lambda(t) = f(b + \vec{k} \cdot \vec{x}(t) + \sum_{j=1}^J h(j)r(t-j)) \quad \text{Eq. (1)}$$

where  $\vec{k} \cdot \vec{x}(t) = \sum_f \sum_\tau k(f, \tau)x(f, t - \tau)$  is the convolution between the stimulus at time  $t$  and the STRF, and  $r(t-j)$  is the cell's spike train history ( $J = 5$ ). The log-likelihood of the observed spike data given the model parameters ( $\vec{\theta} = \{b, \vec{k}, \vec{h}\}$ ), and the observed stimulus  $\vec{x}$  is given by the point process log-likelihood (Snyder and Miller 1991)

$$L = \sum_{t_{spk}} \log(\lambda_{t_{spk}}) - \int \lambda(t) dt \quad \text{Eq. (2)}$$

where  $t_{spk}$  denotes the set of spike times and the integral is taken over the length of the experiment (all trials of the particular stimulus class used to fit the model).

#### 2.4.2 Static nonlinearity

Numerical optimization of Eq. 2 is simplified by making two assumptions about the nonlinear rectification function  $f(\cdot)$ : 1)  $f(u)$  is a convex function of its scalar argument  $u$ ; and 2)  $\log f(u)$  is concave in  $u$ . With these assumptions, the log-likelihood in Eq. 2 is guaranteed to be a concave function of the parameters  $\vec{\theta}$  (Paninski 2004). This ensures that the likelihood has no local maxima, and therefore the maximum likelihood parameters  $\vec{\theta}_{ML}$  may be found by numerical ascent techniques. Several functions  $f(\cdot)$  satisfy these two constraints, including the standard linear rectifier and the exponential function.

For each cell, our model converts linear input into an instantaneous spike rate by means of an exponential nonlinearity (see Fig. 1A). To assess the adequacy of this assumption, we compared an exponential function with a direct reconstruction estimate of the nonlinearity, computed using the raw distribution of filter outputs and the observed spike responses (Chichilnisky 2001) (see Fig. 2C for an example). These reconstructions look exponential for some cells in our sample and sub-exponential for others. To assess the performance of the exponential nonlinearity against another nonlinearity, we also performed a complete re-fitting of the model parameters using output nonlinearities given by a function of the form

$$f = \begin{cases} e^u & u \leq 0 \\ 1 + u + \frac{u^2}{2} & u > 0 \end{cases}$$

which grows quadratically for  $u > 0$  and decays like  $e^u$  for negative values of  $u$ . This model

exhibited slightly better predictive power for ml noise ( $p < 0.05$ , two-sample Kolmogorov-Smirnov test) but not for songs across the population of 169 cells, and did not result in a noticeable change in the fitted STRFs.

The weak dependence of the parameter estimates on the specific form of the nonlinearities tested here (in addition, we also fitted a linear model with a sparse prior which resulted in nearly identical STRFs, see Section 3.6 for further details) led us to ask whether we could improve the performance of the model by fitting a flexible nonlinearity for each cell once the parameters ( $b$ ,  $\vec{k}$ , and  $\vec{h}$ ) were already known (in general, the estimates will depend on the specific form of the objective function used for optimization and a re-estimation step is necessary after the nonlinear function  $f$  is fitted). We parameterized the output nonlinearity as a cubic spline, and used this model instead of the exponential nonlinearity to predict novel responses and compared those to predicted responses that were generated using the exponential nonlinearity. This addition conferred only a slight improvement in cross-validation performance (see Section 3.6 and Fig. 8). Therefore, for simplicity, we restricted all further analyses in this study to a GLM with an exponential nonlinearity.

#### 2.4.3 Regularized sparse solutions

Maximum likelihood estimates can be extremely noisy when fitting high-dimensional models. This overfitting phenomenon has been shown in the linear regression case (Theunissen et al. 2001) (see section 2.5), where the noisiness of the estimate of the filter  $\vec{k}$  is roughly proportional to the dimensionality of  $\vec{k}$  divided by the total number of observed samples (Paninski 2003). The same type of effect occurs in the GLM context. Thus, in order to obtain accurate fits, we added to the log-likelihood in Eq. 2 an additional term,  $Q(\vec{k})$ , that acts as a “penalty function.”

$$L - Q(\vec{k}) = \sum_{t_{spk}} \log(\lambda_{t_{spk}}) - \int \lambda(t) dt - Q(\vec{k}) \quad \text{Eq. (3)}$$

Here  $Q(\vec{k})$  encodes our a priori beliefs about the true underlying  $\vec{k}$ . Whenever the penalizer  $-Q(\vec{k})$  is a concave function of  $\vec{k}$ , the penalized likelihood in Eq. 3 is also a concave function of  $\vec{k}$ , and ascent-based maximization may proceed as before, with no local maxima (Paninski et al. 2004). Thus, the penalty term  $Q$  can be any function within the class of convex functions.

Here we used a sparse prior on the STRF (i.e., many of the elements of  $\vec{k}$  are zero and only a small subset of the elements of  $\vec{k}$  is active) to regularize the model. This is equivalent to assuming that the neuron’s firing is sensitive only to a small number of stimulus features (David et



al. 2007). A common way to impose sparseness is based on the L1 norm of  $\vec{k}$  (Donoho et al. 1995, Tibshirani 1996),

$$Q(\vec{k}) = \eta \sum_i |k_i|. \quad \text{Eq. (4)}$$

This function is convex, but the term on the right in Eq. 4 is non-differentiable and the resulting optimization problem can be challenging. An alternative approach is to use a smooth differentiable approximation to the L1-regularizer that would allow the application of standard Newton methods to solve the resulting unconstrained optimization problem. Within this context, we use the interior point method proposed by (Schmidt et al. 2007) to solve the optimization problem. This method relaxes the non-differentiability of the L1-norm by a sequence of smooth approximation functions. Solving this optimization problem requires the selection of an additional hyperparameter,  $\eta$ , that controls the amount of penalization: for large  $\eta$  we penalize strongly and for  $\eta = 0$  we recover the maximum likelihood unregularized solution (see Fig. 1B). Here, we select this hyperparameter by cross-validation, varying  $\eta$  until a maximum in prediction accuracy is reached.

### 2.5 STRF estimation by normalized reverse correlation

For comparative purposes, we estimated STRFs from the same data using normalized reverse correlation (NRC), a variant of the classical linear regression that has been used to estimate STRFs from natural stimuli in the auditory and visual systems (Theunissen et al. 2000, Theunissen et al. 2001, David et al. 2004). NRC fits a linear STRF that minimizes the mean-squared error between predicted and observed neuronal response:

$$\vec{k}_{LS} = \underset{\vec{k}}{\text{argmax}} \left\{ \sum_t \left[ \vec{k}^T \vec{x}_t - r_t \right]^2 \right\}. \quad \text{Eq. (5)}$$

A detailed description of the algorithm is described in (Theunissen et al. 2001). Here we provide a brief description of NRC for comparison to the GLM method.

The best-fit least-squares filter  $\vec{k}_{LS}$  is given by

$$\vec{k}_{LS} = (X^T X)^{-1} X^T \vec{r} \quad \text{Eq. (6)}$$

where the stimulus matrix  $X$  is defined as  $X_t = x_t^T$  and  $r = (r(1) \ r(2) \ \dots \ r(t))^T$ .

The term  $X^T \vec{r}$  corresponds to the spike-triggered average - the conditional mean  $\vec{x}$  given a

slope - and the matrix  $X^T X/D = C_{XX}$  corresponds to the covariance matrix of the stimulus. Here, the superscript  $T$  indicates a transpose operation and  $D$  is the duration of the experiment.

This estimator gives an unbiased estimate of the filter  $\bar{k}$  for any stimulus statistics if the underlying system is linear (Klein et al. 2000) or if the stimulus is elliptically symmetric (i.e. contains only up to second order correlations) if the underlying system is nonlinear. However, in practice, for the case of high dimensional  $X$  with strong autocorrelations, estimating  $\bar{k}$  with Eq. 6 can amplify noise excessively (Theunissen et al. 2001). To minimize these effects, NRC uses a pseudo inverse to approximate the inverse of the stimulus autocorrelation matrix. This approximation is based on setting dimensions in the stimulus that have little power (below some noise threshold) to zero. To compute the pseudo inverse, a singular value decomposition is applied to the autocorrelation matrix,

$$C_{XX} = U^T \Lambda U \quad \text{Eq. (7)}$$

The columns of  $U$  contain the unit-norm eigenvectors of  $C_{XX}$ , which correspond to the discrete Fourier transform (DFT) vectors. The diagonal matrix  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$  contains the corresponding eigenvalues ordered from largest to smallest, which correspond to the power spectrum of the stimulus as a function of temporal frequency. A tolerance value,  $\tau$ , specifies the fraction of stimulus variance and the number of stimulus dimensions,  $m$ , to preserve in the pseudoinverse  $C_{app}^{-1}$ :

$$C_{app}^{-1} = U^T \Lambda_{app}^{-1} U = U^T \text{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_m}, 0, \dots, 0\right), \quad \text{Eq. (8)}$$

which results in penalization of high frequencies.

The final NRC estimate of the STRF is then,

$$\bar{k} = \frac{1}{D} C_{app}^{-1} X^T \bar{r} \quad \text{Eq. (9)}$$

Implementing NRC requires the selection of a single hyperparameter, the tolerance value,  $\tau$ . To choose  $\tau$ , the method uses a cross-validation procedure. In this study, we use the Matlab toolbox developed by the Theunissen and Gallant laboratories at University of California, Berkeley (<http://strfpak.berkeley.edu>) to estimate NRC STRFs.

### 2.6 *Prediction accuracy*

For both the GLM and NRC models, response prediction was validated with song and noise data that were not used for fitting the model. From the entire set of 20 song and 10 ml noise stimuli, 19 songs or 9 ml noise samples were used to estimate the models' parameters (estimation data set). The models were then used to predict the average response to the remaining song or ml noise sample not included in the estimation set. This procedure was repeated 20 times in the case of songs and 10 times in the case of ml noise, excluding a different validation stimulus on each repeat. The accuracies of these predictions were then averaged to produce a single value for the neuron's response. Prediction accuracy was determined by measuring Pearson's correlation coefficient between the predicted and observed peristimulus time histogram (PSTH). For both NRC and GLM, PSTHs were computed with a 3 ms time bin and were smoothed with a 5 ms Hanning window. The width of the smoothing window was chosen to match the integration time of auditory midbrain neurons (Covey et al. 1996).

### 2.7 *Tuning properties derived from STRFs*

To compare STRFs estimated using NRC and GLM methods, we measured three tuning properties commonly used to characterize auditory neurons (Escabi and Read 2003, Woolley et al. 2006); best excitatory frequency (BF), the spectral frequency that evokes the strongest neural response; excitatory spectral bandwidth (BW), the range of frequencies that are associated with an increase from mean firing rate; and excitatory temporal bandwidth (tBW), the time over which relevant frequencies lead to an increase from mean firing rate. BF, BW and tBW were computed from the STRFs using standard methods (David et al. 2009). Briefly, the BF was measured by setting negative STRF values to zero and averaging along the time axis. The resulting spectral tuning curve was convolved with a 5-point symmetric Hanning window, and the BF was taken to be the position of the peak of the smoothed curve. The BW was measured from the smoothed curve as the width (Hz) at half-height around the BF. The tBW was measured by setting all negative STRF values to zero and averaging along the spectral axis. The resulting temporal tuning curve was convolved with a 5-point symmetric Hanning window, and the tBW was measured from the smoothed curve as the width (ms) at half-height around the peak.

## RESULTS

We recorded from 169 well-isolated MLd neurons (97 in anesthetized birds and 72 in awake birds). Neurons recorded from awake and anesthetized birds produced robust responses to songs and ml noise. On average, midbrain neurons recorded from awake birds showed higher spontaneous and stimulus-driven firing rates, when compared to neurons recorded from

anesthetized birds (mean stimulus-driven firing rates were 22 Hz for the awake preparation and 11 Hz for the anesthetized preparation). At the single neuron level, we did not find significant differences in firing rate in responses to song and ml noise in awake or anesthetized birds. The responses of nearly all neurons were stimulus-locked and reliable over multiple presentations of the same stimulus (trials).

### 3.1 *Responses of single auditory midbrain neurons are well modeled using a GL model*

We modeled the functional relationship between sound stimuli and neuronal responses with a generalized linear model (GLM) for each neuron (see Fig. 1A). Figure 2A shows an example GLM STRF estimated from responses to song and Figure 2B shows the corresponding exponentiated post-spike filter representing the influence of spiking history on spiking probability for the same neuron. For most of the cells in our sample, the shape of the post-spike filter corresponds to a brief period of refractoriness and gradual recovery.

Figure 2C shows the static nonlinearity estimated for this neuron (Chichilnisky 2001) (black dots), together with the exponential nonlinearity (gray line) employed by the model. Although the exponential function used by the model does not provide an excellent fit to the underlying nonlinearity for this neuron (a subexponential nonlinearity performs slightly better; see Methods), the model does predict responses to a novel stimulus with good accuracy (see below).

In order to test how well the GLM method predicted song responses in individual trials, we used it to predict the responses to a validation song (Figure 2, D-F) that was included in the recording experiment but was not included in the estimation of the model parameters. Recorded and predicted spiking responses to the validation stimulus are shown in Figure 2F. For this neuron, the model predicts the spiking responses to the validation song reasonably well; the mean cross correlation between actual and predicted response PSTHs was 0.69.

### 3.2 *A GLM outperforms normalized reverse correlation (NRC) when predicting responses of single auditory neurons to songs and noise.*

We next compared the GLM to the more traditional STRF estimation method, NRC, in their ability to predict single neuron responses to zebra finch songs and ml noise.

Figure 3 shows NRC and GLM response predictions for 3 neurons in response to the song in Figure 2D. Although the predicted traces for both models (blue for NRC and red for GLM) account for broad variations in the actual PSTHs, neither of them captures their precise shape. One common failure of the models to predict responses is best demonstrated in Figure 3B and C.

These two neurons show highly reliable responses to the song and, although the models predict the timing of the responses, in several cases they underestimate their amplitudes. This effect is more pronounced for NRC than for the GLM. Changing the nonlinear link function of the model could, in principle, help to increase the amount of variance in the response described by the model. However, within the groups of nonlinearities we tried on our data (see Methods), we observed only slight or no improvements in prediction accuracy (see also Fig. 8). We later discuss (see Discussion) several extensions to the GLM that could improve the predictive power of the model. Finally, Figure 3 shows spike-train predictions for the GLM method in response to the same song. The predicted spike trains capture the overall structure of the recorded spiking activity.

We then compared the prediction accuracy of the GLM and NRC methods across the entire set of 169 auditory midbrain neurons. Since we did not find noticeable differences in predictive power between awake and anesthetized recordings, we report the prediction accuracy for the two data sets combined.

We first compared the ability of both models to predict responses to a novel stimulus taken from the same stimulus class used in the estimation set (we refer to this case as "same-class predictions", Figure 4A). We found that the performance of both models varies widely across our population of cells; on this moderately small timescale (predicted and actual responses were computed using 3 ms time bins and were smoothed with a 5 ms Hanning window; see Methods), the prediction correlation was as high as 0.77 for some neurons and below 0.1 for others. For low firing rate neurons, we found a relatively moderate correlation (0.36) between the number of spikes in the estimation set and the prediction performance of the models. We found that the prediction performance becomes independent of the number of spikes in the estimation set for  $N \sim 2000$ , which corresponds to firing rates of  $\sim 10$  Hz. Since the goal of this study is to test the GLM method under different conditions and compare its performance to NRC, we included all the data in our sample in the analysis regardless of prediction accuracy.

The average same-class prediction correlation for the GLM for novel song and ml noise stimuli is  $r_s = 0.47 \pm 0.01$  and  $r_n = 0.46 \pm 0.01$ , respectively (mean  $\pm$  SE). These values are significantly greater than the average for NRC,  $r_s = 0.42 \pm 0.01$  and  $r_n = 0.40 \pm 0.008$  ( $p < 0.001$ , two-sample Kolmogorov-Smirnov (KS) test).

To evaluate how well the GLM and NRC methods estimated from responses to one of the stimulus classes generalized to a second stimulus domain, we compared how well these models predicted responses to the other stimulus class ("across-class predictions"). In this way, we used the models that were estimated using song data to predict responses to ml noise and vice-versa (Figure 4B). As in the case of same-class predictions, the GLM predicts responses to the

opposite class ( $r_s = 0.38 \pm 0.01$  and  $r_n = 0.4 \pm 0.01$ ) significantly better than NRC ( $r_s = 0.34 \pm 0.01$  and  $r_n = 0.29 \pm 0.01$ ,  $p < 0.001$ , two-sample KS test).

The absolute prediction accuracy for both models in the across-class case is significantly lower than in the same-class case. For the GLM, the mean prediction correlation is 15% lower in the across-class case than in the same-class case, both for noise and song predictions. For NRC, the mean prediction correlation is 15% lower in the across-class case than in the same-class case for song predictions, and 28% lower for noise predictions. This decrease in performance suggests that neither model generalizes completely to other stimulus classes. Because of nonlinear response properties, STRFs estimated using one stimulus class tend to predict responses to other stimulus classes with worse accuracy (Woolley et al. 2006, Sharpee et al. 2008). However, the better performance of the GLM suggests that it provides a more general characterization of spectrotemporal tuning across different stimulus conditions.

### 3.3 *GLM STRFs are more stable to changes in the stimulus statistics than are NRC STRFs.*

The fact that the GLM produces better response predictions across stimulus classes than does NRC (see Figure 4B), suggests that it generalizes better to changes in the statistics of the stimulus used to estimate the model. In agreement with this, we found that GLM STRFs were more similar to each other between stimulus classes than NRC STRFs for the entire population of 169 cells. Figure 5, A-C shows GLM (top panel) and NRC (bottom panel) STRFs derived from responses to song (Ks) or ml noise (Kn) for three example neurons. In agreement with previous observations (Woolley et al. 2006, Sharpee et al. 2008, David et al. 2004), we found that, for some neurons, NRC STRFs estimated from different stimulus classes show significant differences (see, for example Figure 5B and 5C). Figure 5A shows an example neuron for which Kn and Ks do not differ, and Figure 5, B-C shows example cells for which NRC STRFs estimated from recorded responses to ml noise and song differ significantly. In contrast, GLM song and noise STRFs appear significantly more similar for all three cells.

To quantify the differences between song and ml noise STRFs (Kn and Ks) derived using a GLM or NRC, we measured a similarity index (SI, pixel by pixel correlation). A similarity index of 1 indicates a perfect match, and an index of 0 indicates no correlation between STRFs. Figure 5D shows the distributions of SIs between Kn and Ks for the GLM (white) and NRC (grey) for the set of 169 neurons. These distributions differ significantly ( $p < 0.001$  KS test). The shift to the right in the SI distribution for the GLM shows that GLM STRFs are more similar across stimulus classes than are NRC STRFs at the population level (median of the SI distribution 0.76 for the GLM and 0.61 for NRC).

In addition, we found that differences between NRC and GLM STRFs derived from responses to ml noise (compare GLM Kn vs. NRC Kn in Fig. 5, A-C) were smaller than differences between NRC and GLM STRFs derived from responses to songs (compare GLM Ks vs. NRC Ks in Fig. 5, A-C), as would be predicted theoretically. When stimuli that contain only second-order correlations are used to derive the STRF, NRC and GLM should give the same answer in the limit of infinite data (Chichilniski 2001, Paninski 2003, Sharpee et al. 2004). Non-Gaussian effects in the ml noise stimulus ensemble are smaller than in the song ensemble, which explains the smaller differences between GLM and NRC STRFs for this stimulus class.

### 3.4 *Tuning properties of GLM and NRC STRFs*

In Section 3.3 a nonparametric comparison between song and ml noise STRFs derived under a GLM or NRC showed that GLM STRFs are more similar across these two stimulus classes than are NRC STRFs. Measures of excitatory tuning taken from STRFs showed significant differences between NRC and GLM STRFs (Figure 6). Best frequency (BF) did not differ between NRC and GLM STRFs (Figure 6A). Excitatory spectral bandwidths (BW) were significantly different between NRC and GLM STRFs; NRC BWs were larger than GLM BWs (Figure 6B). The mean BW for song STRFs was  $1312 \pm 100$  Hz for NRC and  $917 \pm 41$  Hz GLM, and  $703 \pm 31$  and  $798 \pm 35$  for noise STRFs. For both song and ml noise, the difference in BW between NRC and GLM STRFs was highly significant ( $p < 10^{-3}$ , two-sample KS test). Excitatory temporal bandwidths (tBW) also differed between NRC and GLM STRFs (Figure 6C). For song STRFs, the mean tBW was  $9.4 \pm 0.5$  ms for NRC and  $5.6 \pm 0.16$  ms for GLM. For ml noise STRFs, mean tBWs were  $7.9 \pm 0.17$  ms and  $5.1 \pm 0.12$  ms, respectively. For both song and ml noise, the difference in tBW between NRC and GLM STRFs was highly significant ( $p < 10^{-3}$ , two-sample KS test).

When comparing song and noise STRFs within a neuron, we found no significant differences in BF for the NRC or GLM (Figure 6D,  $p > 0.9$ , two-sample KS test). The differences in BW between song and noise STRFs were significantly larger for NRC than for GLM (Figure 6E,  $p < 10^{-3}$ , two-sample KS test). Finally, we found no significant differences in tBW between noise and song GLM STRFs (Figure 6F,  $p > 0.1$ ), but differences were significant for NRC STRFs (Figure 6F,  $p < 10^{-3}$ ).

In summary, for the population of neurons studied here, GLM and NRC STRFs estimated from the same song and ml noise data show substantial differences. Further, ml noise and song STRFs differed significantly in their spectral and temporal properties, but differences were larger for NRC STRFs than for GLM STRFs.

### 3.5 *Effects of estimation algorithm-induced biases on STRFs*

As mentioned earlier, for a linear neuron, reverse correlation (RC) methods are guaranteed to produce an unbiased estimate of a neuron's true underlying STRF regardless of the stimulus statistics (Klein 2000). For a linear-nonlinear (LN) neuron, RC is guaranteed to produce an unbiased estimate of a neuron's true underlying filter only if the distribution of the stimuli used for estimation is elliptically symmetric (Paninski2003). However, in the presence of stimuli with higher-order correlations, such as zebra finch songs, RC can introduce biases in the estimate of the STRF. Something similar occurs with the GLM: if the underlying neuron behaves like a GLM, then a GLM will produce an asymptotically unbiased estimate of the STRF of the cell for any stimulus ensemble. However, any deviation from the GLM framework can introduce biases in the estimates (Paninski 2004).

In addition, the highly correlated structure of zebra finch songs presents additional numerical problems for STRF estimation, causing noise in the resulting STRF to be strongly amplified (see Methods for further details). Thus, some form of regularization is applied to the estimation method to obtain accurate STRFs (Theunissen et al. 2001, Sahani and Linden 2003, Smyth et al. 2003, Machens et al. 2003). In the presence of limited or noisy data (a common scenario in neurophysiological experiments), regularization introduces a prior that constrains the STRF estimate in a way that is independent of the underlying tuning properties of the neuron, but can introduce additional biases in the STRF. Because of these types of effects, in some cases, STRFs can reflect statistical properties of the stimuli used for estimation or biases introduced by the estimation algorithm (e.g. the particular prior) rather than actual tuning properties of the underlying neuron (Christianson et al. 2008, David et. al 2007).

We asked whether and how much of the tuning differences we observe between song and ml noise STRFs (see Figs. 5 and 6) can be explained in terms of biases introduced by the estimation algorithm. To address this, we used Kn and Ks (see Section 3.3) as LNP-type generative models to synthesize responses to both stimulus classes and re-estimate the STRFs.

Briefly, for NRC, we generated synthetic responses to song or ml noise with the following model:  $r_s = k \cdot x_s + b$  or  $r_n = k \cdot x_n + b$ , respectively. Similarly, for the GLM, we generated responses to both stimulus classes using Eq. 1. Here  $K$  and  $b$  (and  $h$ , for the case of the GLM) were either derived from recorded responses to song ( $K_s$ ,  $b_s$ ) or ml noise ( $K_n$ ,  $b_n$ ). Thus, we are left with two types of synthetic responses to song ( $r_{ss}$  and  $r_{ns}$ ), and two types of synthetic responses to noise ( $r_{sn}$  and  $r_{nn}$ ), which correspond to using  $K_s$  or  $K_n$  in the generative model. These four sets of responses were then used to compute two second-generation ml noise STRFs ( $K_{nn}$  and  $K_{ns}$ ) and two second-generation song STRFs ( $K_{sn}$  and  $K_{ss}$ ) derived from synthetic responses to ml noise or songs, respectively. The differences between these new STRFs and the original STRFs were then quantified. Our rationale was that, if the estimation algorithms were free of biases, we should recover  $K_n$  and  $K_s$  with some added noise, regardless of the stimulus class used to re-



estimate the STRFs. In particular,  $K_{nn}$  and  $K_{ns}$  should show small differences when compared to  $K_n$ , and  $K_{sn}$  and  $K_{ss}$  should show small differences when compared to  $K_s$ .

Figure 7, A-C shows the original NRC STRFs derived from recorded data and the re-estimated STRFs for the same three cells shown in Fig. 5. We found that, in some cases, the underlying noise and song STRFs are recovered by the simulations (Figure 7A, compare  $K_n$  with  $K_{ns}$  and  $K_{nn}$ , and  $K_s$  with  $K_{sn}$  and  $K_{ss}$ ). However, we also found cases for which the simulated noise and song STRFs differ significantly from the ones derived from recorded responses (Figure 7, B-C). In the example shown in Figure 7B, biases in the estimation algorithm are not sufficient to explain the original differences between  $K_n$  and  $K_s$ , indicating the presence of actual nonlinearities in the responses that result in stimulus-dependent tuning. In contrast, for the example shown in Figure 7C, differences between  $K_n$  and  $K_s$  can be explained by biases introduced by the estimation algorithm (that is,  $K_{ns}$  is significantly more similar to  $K_s$  than to  $K_n$ , even though the responses used to compute  $K_{ns}$  were originally generated from  $K_n$ ).

We next repeated the same type of simulations for the GLM. Figure 7,D-F shows the results of these simulations for the same three cells in Figures 7, A-C and 5. Figure 7D shows an additional example cell with broader spectral tuning and stronger inhibition. For all four examples, the GLM recovers STRFs from the simulations that are highly similar to the true underlying STRFs. Thus in these examples, it is visually clear that the GLM reduces algorithm-induced biases in the STRFs.

To quantify the amount of biases introduced in the STRFs by the GLM and NRC at the population level, we measured the similarity index (SI) between STRFs derived from recorded and simulated data. In particular, we measured the similarity between  $K_n$  and  $K_{ns}$  (STRF estimated from synthetic responses to song when the true underlying filter in the LNP-model was  $K_n$ ), and between  $K_s$  and  $K_{sn}$  (STRF estimated from synthetic responses to ml noise when the true underlying filter in the LNP-model was  $K_s$ ). Figure 7H shows these distributions for the GLM (white) and NRC (grey) for our set of 169 cells. If the estimation algorithms introduced no (or little) bias in the STRF estimates, then the SI distribution should be, up to some variability, a narrow distribution located close to 1. For NRC, we observe a broad distribution with median = 0.64 when we measure SIs between  $K_n$  and  $K_{ns}$ , and with median = 0.73 when we measure SIs between  $K_s$  and  $K_{sn}$ . In contrast, SI distributions for the GLM are narrower and centered closer to 1 (median = 0.94 and 0.87, respectively), and differ significantly from NRC distributions ( $p < 0.001$ , KS test, in both cases).

These analyses show that, in some cases, differences in tuning between STRFs derived from responses to song and ml noise stimuli can be explained in terms of biases introduced by the estimation algorithm, rather than actual tuning nonlinearities (Christianson et al., 2008). However, these effects are exaggerated when NRC instead of the GLM is used.

### 3.6 *Effect of the regularization prior on STRFs and predictive power.*

In Section 3.2 we showed that the GLM has a higher predictive power than NRC, both within and across stimulus classes (see Figure 4). In addition, and in accordance with a higher across-class prediction power, we found that the GLM produces STRFs that are more similar across stimulus classes than does NRC (Figures 5 and 6). We also found that, in some cases, the differences between song and noise NRC STRFs as well as the differences between NRC and GLM STRFs can be explained in terms of biases introduced in the STRFs by NRC (See Figure 7). In this Section, we address a related question: what component of the GLM is responsible for reducing algorithm-induced biases in the STRFs and at the same time increasing the predictive power of the model?

The GLM and NRC methods differ in three ways that result from the different assumptions about neural responses made by each of the two methods. First, the two methods optimize different objective functions; the GLM assumes point-process responses with an exponential nonlinearity while NRC assumes Gaussian noise and uses a simpler linear model. Second, the GLM and NRC use different regularization methods. Our method imposes a sparse prior on the STRF while NRC uses a lowpass Gaussian prior. Third, the GLM includes a spike history term that takes into account the recent firing probability of the neuron, while NRC does not. In principle, each of these factors may contribute to the better predictions and less-biased STRFs produced by the GLM. In order to study the effect of each component of the GL model, we remove each of these factors from the GLM framework.

We first tested the hypothesis that the differences observed between NRC and GLM STRFs, and the higher predictive power of the GLM, are due to the fact that our method optimizes a different objective function than NRC. In particular, the nonlinearity employed by the GLM might be the important difference between the models. To test this, we re-fitted the GLM with a sub-exponential nonlinearity that was closer to the actual response nonlinearity in the data (see Methods). We found that this led to a slight (but statistically not significant) improvement in the predictive power and, importantly, no change in the shape of the STRFs. This weak dependence of the STRFs on the specific nonlinearity led us to ask whether it was possible to increase the predictive power of the model by fitting a cubic spline nonlinearity for each cell once the model parameters were already known (we refer to this model as spline GLM, see Methods for further details). This flexible nonlinearity conferred only a slight (but not significant) increase in predictive power for songs but not for ml noise responses when compared to the exponential GLM (see Figure 8).

Another difference between the two models is the extra term in the GLM that accounts for the neuron's spiking history. The post-spike filter could contribute to changes in the prediction accuracy of the model and in the shape of the STRF. To account for the effects of the post-spike filter on predicted responses, we re-fitted the GLM without this term, referred to here as exponential LN. We found no differences between GLM and LNP STRFs. We did, however, find that the prediction power of the GLM was slightly (but not significantly) better than the prediction power of the LN model when trying to predict responses to noise (Figure 8). It is likely that the small contribution of the spiking history to response prediction is due to the relatively low firing rate of the neurons in our sample. Spiking history may contribute to an accurate description of the detailed structure of response spike trains in higher firing rate neurons, as has been shown in the retina (Pillow et al. 2005).

Finally, to test the effect of the regularization prior on predictive power and STRF shape, we fitted the same linear model used by NRC (Eq. 5), but with an L1 regularizer (Eq. 4) instead of the lowpass Gaussian prior imposed by NRC (we refer to this model as linear L1). We found that the linear L1 model has significantly higher predictive power than does NRC (see Figure 8). We found no significant difference in predictive power between the linear L1 model and the nonlinear models, with or without the history term (i.e., Exp GLM, spline GLM and Exp LN). Importantly, we found no significant differences in STRF shapes. These comparisons between STRFs and predictions generated by different models employing an L1 regularizer and NRC indicate that the differences between NRC and GLM are mostly due to the fact that the two estimation algorithms assume different priors about the STRF.

## DISCUSSION

We used a generalized linear model (GLM) with a sparse prior to characterize the stimulus-response relationships of single auditory midbrain neurons, and compared the performance of our model to that of normalized reverse correlation (NRC) for predicting the responses to novel sounds. We found that a GLM can be successfully used to predict single-trial responses to synthetic and natural stimuli, and that, for the population of 169 cells used in this study, the GLM had a better predictive power than NRC. The performance of the GLM was better than NRC both within and across stimulus classes. The good performance of the GLM across stimulus classes suggests that our method generalizes better to changes in stimulus statistics. Differences between STRFs computed from responses to different stimulus classes (e.g. song and noise STRFs) were significantly smaller than those observed when STRFs were computed with NRC. Differences in the STRFs computed with the GLM and NRC methods were largely due to differences in the estimates of excitatory spectral bandwidths and temporal bandwidths. Below,

we discuss the computational differences between the GLM and NRC that lead to differences in predictive power and STRF shapes, and compare the GLM method to other proposed methods for characterizing stimulus-response relationships in auditory neurons.

#### *Computational differences between the GLM with a sparse prior and NRC*

As discussed before (see Section 3.6), the sparse GLM and NRC contain three fundamental differences: the two methods optimize different objective functions, use different regularization methods, and the GLM takes into account the recent firing probability of the neuron, while NRC does not. For the neurons studied here, we found that the differences in STRFs produced by both methods and the higher predictive power of the GLM are largely due to the different priors used by the two methods. Even though the GLM takes into account the cell's spiking history and uses a different nonlinearity than NRC, we found that the contribution of the spike history term as well as more complex nonlinearities led to little or no increase in the model's predictive power (see Methods and Section 3.6), and no noticeable change in STRFs' shape.

NRC estimates the STRF only in the stimulus subspace that contains most of the variance of the stimulus to reduce noise in the estimates and avoid overfitting (Theunissen et al. 2001). With increasing levels of noise, and depending on the specific spectrotemporal characteristics of the stimulus, NRC produces STRFs that are biased towards being smooth (see Methods). This is particularly the case for songs and other natural stimuli for which the majority of the power tends to be concentrated at low spectrotemporal frequencies (Singh 2003). In this case, the spectral and temporal features at high frequencies tend to be excluded from the STRFs estimated using NRC, resulting in STRFs that substantially overestimate the contribution of low-frequency components to neural filtering (David et al. 2007, Sharpee et al. 2008). In contrast, the sparse GLM imposes a sparse prior on the STRFs. In this case, the amount of regularization applied to the STRF depends on the overall level of noise in the data, and in the case of a low signal-to-noise ratio, GLM STRFs will be overly sparse (see Methods). Even though both NRC and GLM methods introduce biases in the STRFs, the GLM leads to better predictions and model stability across stimulus classes.

#### *Comparison of the GLM to other methods*

An alternative approach for estimating a sparse STRF is boosting (Zhang and Yu 2005, Friedman et al. 2008). Boosting is an estimation technique that uses coordinate ascent to minimize the number of nonzero parameters, effectively imposing a sparse prior on the STRF. David et al. (2007) applied boosting on the same objective function as NRC (i.e. a linear model), to derive

STRFs for primary auditory cortex neurons. Their results showed, in agreement with our findings, that boosting STRFs lead to better prediction power and show narrower spectral and temporal bandwidths than do NRC STRFs. The differences in predictive power between NRC and GLM reported here are slightly larger than the differences reported in (David et al. 2007) for NRC and boosting STRFs. This is presumably due to the different nonlinearities employed by the two methods. Finally, boosting can also be applied to estimate GLMs with L1 penalties (Friedman et al. 2008).

Several other algorithms have been developed for STRF estimation in the visual and auditory systems (Ringach et al. 2002, Sahani and Linden 2003, Sharpee et al. 2004). Maximally informative dimensions (MID) (Sharpee et al. 2004) is an information-theoretic method that finds relevant directions (a set of  $\vec{k}$  vectors,  $\{\vec{k}_1, \vec{k}_2, \dots\}$ ) in the stimulus space. In its one dimensional version (1d-MID), this method searches for the spectrotemporal filter or STRF  $\vec{k}$ , whose output,  $\vec{k} \cdot \vec{X}$ , carries the most mutual information about the measured neural response  $r(t)$ . Once the filter is known, the nonlinearity of the LNP model is computed from the recorded data. With the GLM method, we first find the filter  $\vec{k}_1$  for a fixed nonlinearity (e.g. an exponential function) by maximizing the corresponding likelihood, and then use the filter to fit the output nonlinearity to the recorded data (see Methods). It has been shown that in a number of problems, including estimation of GLMs, maximizing information is equivalent to performing likelihood maximization (Kouh and Sharpee 2009). Thus, if in the GLM method, we iterate between estimating the STRF  $\vec{k}_1$  for a fixed nonlinearity and fitting the nonlinearity of the model to the recorded data, the 1d-MID and GLM methods are almost equivalent (almost because in the way the MID method is usually employed, it does not contain any spike history terms). Here, however, we have shown that for our data set, the estimated filter is only weakly dependent on the specific form of the nonlinearity (see Methods and Section 3.6), which makes the iteration procedure in the GLM unnecessary.

Another useful method for STRF estimation is evidence optimization, introduced by (Sahani and Linden 2003). This method uses a Bayesian approach to include both sparse and smooth “optimized priors” on the STRFs. These prior distributions are optimized with reference to the data, and thus they are no longer priors in the strict sense and instead become part of a hierarchical probabilistic model. The authors show that, by learning hyperparameters that control the smoothness and sparsity of the STRF in a linear model, it is possible to improve the predicting power of a model that considers only sparseness or smoothness of the estimates.

Finally, an appealing future research direction is known in the statistics literature as Bayesian LASSO (Park and Casella 2008). This method is potentially advantageous because it provides

Bayesian error bars for the estimates, and is based on integrating over the posterior distribution instead of maximizing it and has some advantages in terms of how much sparsity can be enforced. This method has been previously applied (Park and Casella 2008, Carvalho et al. 2009) to L1-linear regression problems but this can be easily generalized for GLMs.

### *Extensions of the GLM*

The same approach used by (Sahani and Linden 2003) to combine smoothness and sparsity priors in a linear model can be applied to a GLM. Because both smooth and sparse regularization frameworks have been shown to improve the prediction power of unregularized models (Theunissen et al. 2001, Sharpee et al. 2008, Sahani and Linden 2003), it is likely that combining features of both methods can further improve the quality of the estimates. For instance, using a prior that combines smoothness and sparsity would allow recovering smooth STRFs, while suppressing the apparent background estimation noise at high spectrotemporal frequencies.

It has been shown that the prediction performance of an LN model can be increased by using a nonlinear transformation of the stimulus (e.g., a transformation may capture nonlinearities at earlier stages of processing) that precedes the linear filtering stage (Gill et al. 2006, Ahrens et al. 2008a, Ahrens et al. 2008b). (Gill et al. 2006) showed that a transformation of the sound stimulus using a biologically inspired model of the first stages of auditory processing (Lyon1982) prior to STRF estimation with NRC led to better predictions. This model incorporates the approximate logarithmic spacing of filter center frequencies (log at high frequencies and more linear at low frequencies) in the auditory nerve and an adaptive gain control mechanism, which was important for improving the predictive power of the model. In a different approach, (Ahrens et al. 2008a) utilizes a learned nonlinear transform on the stimulus that converts the initial numerical representation of a stimulus value to a new representation that provides optimal input to the subsequent model. The authors apply this technique to fit an LN model to data from rodent barrel cortex, and showed that the model predicts responses to novel data accurately. Both of these two approaches can be easily applied when fitting a GLM to auditory data.

Two applications of the GLM setting are fast optimal stimulus decoding (Ahmadian et al. 2009), and optimal stimulus design (Lewi et al. 2009). Stimulus reconstruction methods provide an important tool for understanding how sensory information is represented in neural activity. For high-dimensional stimuli such as sound spectrograms, an encoding model that suitably describes how stimuli are transformed into the spike trains of a neuron constitutes a key component for developing efficient decoding methods (Mesgarani et al. 2009, Ramirez et al. 2010). Adaptive experimental designs, on the other hand, are particularly valuable in domains where the data are expensive or limited. This is particularly the case in STRF estimation, which requires the

exploration of high-dimensional stimulus spaces, and where the inability to collect enough data has important consequences on the estimates. The GLM method described here permits the development of efficient algorithms for optimally adapting the experimental design, allowing more efficient data collection (Lewi et al. 2009).

#### ACKNOWLEDGMENTS

We thank Alex Ramirez for his helpful comments.

#### GRANTS

This work was supported by The Gatsby Initiative in Brain Circuitry, NIH training grant #HD007430-11 and NIH NRSA grant DC010301-01A110.

#### REFERENCES

Aersten AM, Johannesma PI (1981) A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol Cybern* 42: 145-156.

Ahmadian Y, Pillow J, Paninski L (2009) Efficient Markov Chain Monte Carlo methods for decoding population spike trains. COSYNE 2009, Salt Lake City, Utah.

Ahrens MB, Paninski L, Sahani M (2008) Inferring input nonlinearities in neural encoding models. *Network: Comput Neural Sys* 19: 35-67.

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.

Brillinger D (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics* 59: 189-200.

Carvalho CM, Polson NG, Scott JG (2009) The horseshoe estimator for sparse signals (<http://ftp.isds.duke.edu/WorkingPapers/08-31.html>).

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network: Comput Neural Sys* 12: 199-213.

Christianson GB, Sahani M, Linden JF (2008) The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J Neurosci* 28:446-455.

Covey E, Kauer JA, Casseday JH (1996) Whole-Cell patch clamp recording reveals subthreshold sound-evoked postsynaptic currents in the inferior colliculus of awake bats. *J Neurosci* 16: 3009-

3018.

David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of V1 neurons. *J Neurosci* 24: 6991–7006.

David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Comput Neural Sys* 18: 191-212.

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.

Donoho DL, Johnstone IM, Kerkycharian G, Picard D (1995) Wavelet shrinkage: Asymptopia? *J R Statist Soc B*, 57: 301-337.

Eldawlatly S, Jin R, Oweiss KG (2009) Identifying functional connectivity in large-scale neural ensemble recordings: A multiscale data mining approach. *Neural Comput* 21: 450-477.

Escabi MA, Read HL (2003) Representation of spectrotemporal sound information in the ascending auditory pathway. *Biol Cybern* 89: 350-362.

Friedman JH, Hastie TJ, Tibshirani RJ (2008) Regularization paths for generalized linear models via coordinate descent (<http://www-stat.stanford.edu/~hastie/Papers/glmnet.pdf>).

Gill P, Zhang J, Woolley SMN, Fremouw T, Theunissen FE (2006) Sound representation methods for spectro-temporal receptive field estimation. *J Comput Neurosci* 21: 5-20.

Kay S (1993) *Fundamentals of Statistical Signal Processing: estimation theory vol 1* Upper Saddle River, NJ: Prentice-Hall.

Klein DJ, Simon JZ, Depireux DA, Shamma SA (2000) Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9: 85-111.

Kouh M, Sharpee TO (2009) Estimating linear-nonlinear models using Renyi divergences. *Network: Comput Neural Sys* 20: 49-69.

Lewi J, Butera R, Paninski L (2009) Sequential optimal design of neurophysiology experiments. *Neural Comput* 21: 619-687.

Lyon RF (1982) A computational model of filtering, detection, and compression in the cochlea. In: *IEEE Int Conf Acoust, Speech and signal processing*, Paris, IEEE, France.



Machens C, Wehr M, Zador A (2003) Spectro-temporal receptive fields of subthreshold responses in auditory cortex. *NIPS* 15: 149-156.

Mesgarani M, 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.

Paninski L (2003) Convergence properties of some spike-triggered analysis techniques. *Network: Comput Neural Sys* 14: 437-464.

Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Comput Neural Sys* 15: 243-262.

Paninski L, Pillow J, Simoncelli E (2004) Maximum likelihood estimation of a stochastic integrate-and-fire neural model. *Network: Comput Neural Sys* 15: 243-262.

Paninski, L, Pillow, J, Lewi, J (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. In: *Computational Neuroscience: Progress in Brain Research*, edited by Cisek, P, Drew T and Kalaska, J, Elsevier.

Park T, Casella G (2008) The Bayesian LASSO. *J Am Stat Assoc* 103: 681-686.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005). Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. *J Neurosci* 25: 11003-11013.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008). Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature* 454: 995-999.

Ramirez A, Ahmadian Y, Schumacher J, Schneider D, Woolley SMN, and Paninski L (2010). Incorporating Naturalistic Correlation Structure Improves Spectrogram Reconstruction From Neuronal Activity in the Songbird Auditory Midbrain. *J. Neurosci*, under review.

Ringach DL, Hawken MJ, Shapley R (2002) Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J Vis* 3: 12-24.

Sahani M, Linden JF (2003) Evidence optimization techniques for estimating stimulus-response functions. *NIPS* 15: 317-324.

Schmidt M, Fung G, Rosales R (2007) Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches. European Conference on Machine Learning (ECML).

Schneider DM, Woolley SMN (2010) Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *J Neurophysiol*. In press.

Sharpee TO, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16(2): 223-250.

Sharpee TO, Miller KD, Stryker MP (2008) On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. *J Neurophysiol* 99: 2496-2509.

Simoncelli E, Paninski L, Pillow J, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. In: *The cognitive neurosciences*. Cambridge: MIT Press.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 116: 3394-3411.

Smyth D, Willmore B, Baker G, Thompson I, Tolhurst D (2003) The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J neurosci* 23: 4746-4759.

Snyder D, and Miller M (1991) *Random Point processes in time and space*. Springer-Verlag.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.

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: Comput Neural Sys* 12: 289-316.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Statist Soc B*, 58: 267-288.

Truccolo W, Eden UW, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93: 1074-1089.

Woolley SMN, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8: 1317-1379.

Woolley SMN, Gill P, and Theunissen FE (2006) Stimulus-dependent auditory tuning results in synchronized population coding of vocalizations in the songbird midbrain. *J Neurosci*, 26: 2499-2512.

Zevin JD, Seidenberg MS, Bottjer SW (2004) Limits on reacquisition of song in adult zebra finches exposed to white noise. *J Neurosci* 24: 5849-5862.

Zhang T and Yu B (2005) Boosting with early stopping: convergence and consistency. *The Annals of Statistics* 33: 1538-1579.

Zhao M, Iyengar S (2010) Nonconvergence in logistic and poisson models for neural spiking. *Neural Comput* 22: 1-14.

## Figure legends

**Figure 1: Methods.** (A) Generalized linear model (GLM) schematic. Each neuron has a stimulus filter or STRF ( $k$ ), and a post-spike filter ( $h$ ) that captures dependencies on the cell's own spiking history. Summed filter output passes through a static nonlinearity  $f$  to produce the instantaneous spike rate. (B) Illustration of the effect of a sparse prior on the STRF estimate. Panels from left to right show STRFs estimated by maximum penalized likelihood for increasing values of the penalization parameter  $\eta$ . Low values of  $\eta$  lead to noisy estimates. For very high values of  $\eta$ , very few parameters are nonzero. The optimal value of  $\eta$  is determined by cross-validation (STRFs have been plotted in their raw, low resolution state).

**Figure 2: A GLM as a neural encoding model.** (A)-(B) Estimated parameters for an example auditory midbrain neuron. (A) STRF. (B) Exponentiated post-spike filter, which may be interpreted as a spike-induced gain adjustment of the neuron's firing rate. It produces a brief refractory period and gradual recovery (with a slight overshoot). (C) Estimate of the nonlinearity transforming linear input to instantaneous spike rate (black points), for the same example neuron (Chichilnisky2001). The nonlinearity represents the probability of observing a spike for each value of net linear input ( $b+k*x+h*r$ ). An exponential function (grey line), the assumed nonlinearity for the model, provides a reasonable approximation to this function. (D) Spectrogram ( $x$ ) of one example song used in the experiments. (E) Stimulus filtered by STRF,  $k*x$ . (F) Recorded (gray) and predicted (red) raster plots in response to the validation stimulus shown in (D).

**Figure 3: The GLM outperforms NRC when predicting responses to songs.** (A)-(C) Examples of NRC and GLM response prediction (PSTHs and spike trains) and corresponding GLM STRFs for three auditory midbrain neurons. Recorded and predicted responses correspond to the song shown in Figure 2D. Spike trains and PSTHs were computed with a 3 ms time bin and PSTHs were smoothed with a 5 ms Hanning window prior to computing correlation coefficients. (STRFs have been up sampled by a factor of 3 for visualization). The GL model performs consistently better than NRC when used to predict average responses to a validation stimulus. In addition, the GLM spike-train predictions capture the overall structure of the actual spike trains.

**Figure 4: The GLM has higher predictive power than does NRC within- and across stimulus classes.** Both methods were evaluated by their ability to predict responses to a validation song and ml noise data set that was not used for parameters' estimation. (A) Same-class predictions: NRC and GLM were used to predict responses to a novel song or ml-noise stimulus when only songs or ml noise were used to train the model. (B) Across-class predictions: NRC and GLM were used to predict responses to a novel song or ml noise stimulus when the other stimulus ensemble was used to train the model. Each point plots the correlation coefficient between the observed and predicted average response (PSTH), for NRC (horizontal axis) and GLM (vertical

axis) for a single neuron. White dots indicate responses to ml-noise and gray dots indicate responses to song. We found that on average, the GLM predicts responses significantly better than NRC ( $*p < 0.001$ , two-sample KS test) both in the same-class and across-class cases. Error bars represent SEs.

**Figure 5: GLM STRFs are more similar across stimulus classes than are NRC STRFs. (A)-(C)** Top row: GLM STRFs estimated from recorded responses to ml-noise (Kn) and song (Ks) for three example midbrain auditory neurons. Bottom row: ml-noise (Kn) and song (Ks) NRC STRFs for the same three cells. **(A)** An example cell for which Kn and Ks do not differ, both for the GLM and NRC. **(B)-(C)** Two example cells for which Kn and Ks differ significantly for NRC, but differ much less for the GLM. **(D)** Distributions of similarity indices (SIs) computed between Kn and Ks for NRC (grey) and GLM (white) for the population of 169 cells. One indicates an exact match between STRFs and 0 indicates no correlation. The GLM produces STRFs that are more similar across stimulus classes as seen by the shift to the right of the SI distribution (median of the SI distribution for the GLM was 0.76 as opposed to 0.61 for NRC). GLM and NRC distributions differ significantly ( $p < 0.001$ , two-sample KS test).

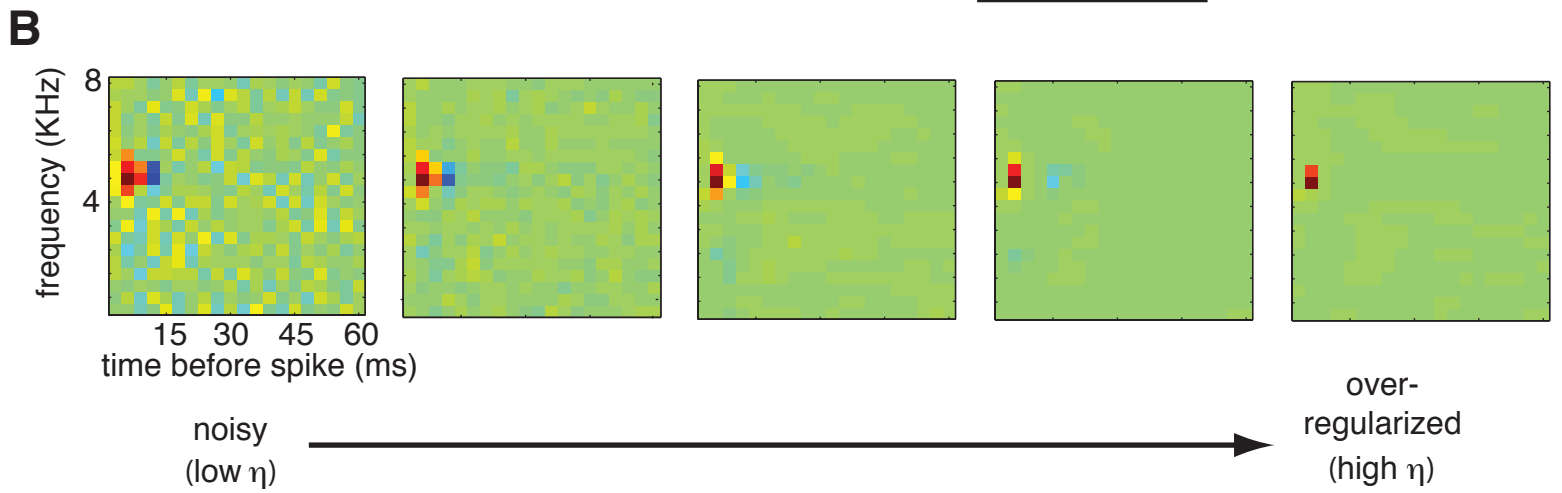
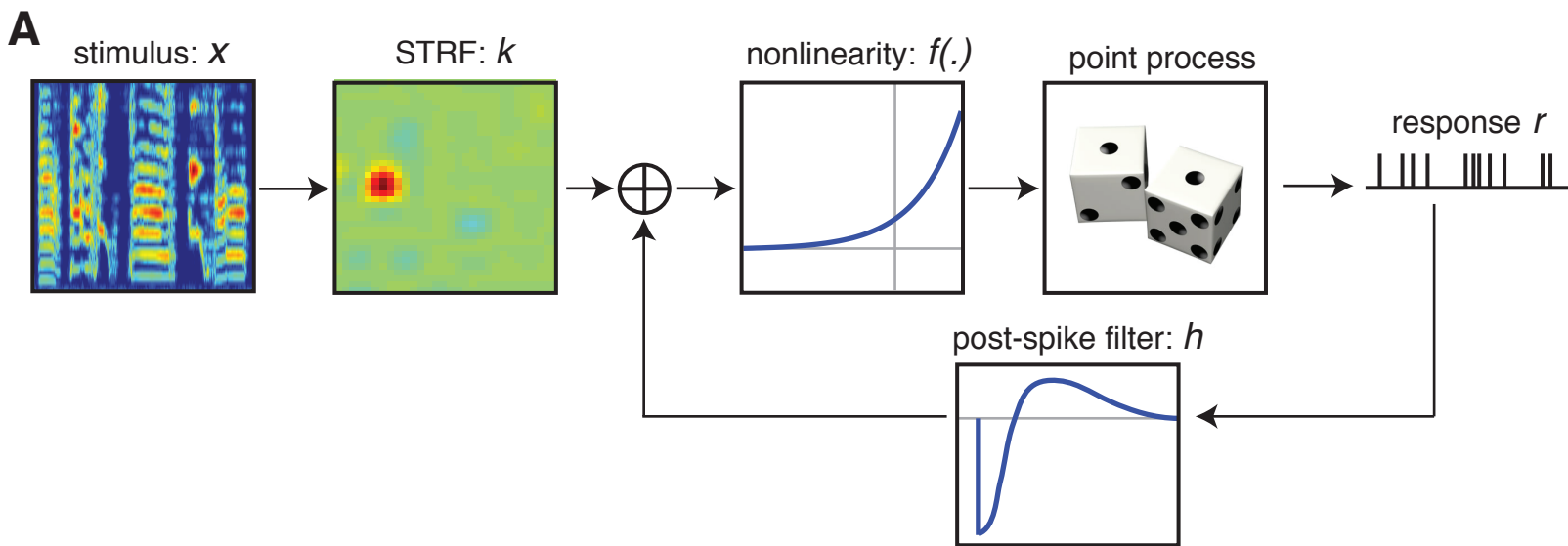
**Figure 6: Tuning differences between GLM and NRC STRFs. (A)-(C)** Comparison of tuning properties of GLM and NRC STRFs. **(A)** Best frequency of the excitatory region (BF), **(B)** excitatory spectral bandwidth (BW), and **(C)** excitatory temporal bandwidth (tBW). We found no significant differences in BF between NRC and GLM STRFs derived from neural responses to ml noise, or those derived from responses to song ( $p > 0.9$ , two-sample KS test). However, we found that differences in BW and tBW determined by the estimation algorithms were highly significant ( $**p < 10^{-3}$ , two-sample KS test). **(D)-(E)** Comparison of tuning properties of song and noise STRFs. **(D)** BF, **(E)** BW and **(F)** tBW. We found no significant difference in BF between song and noise STRFs in either of the models ( $p > 0.9$ ). Differences in BW between song and noise STRFs were considerably larger for NRC than for GLM ( $**p < 10^{-3}$  and  $*p < 0.05$ ). Finally, we found significant differences between song and noise STRFs in terms of tBW for NRC but not for the GLM ( $*p < 0.05$  and  $p > 0.1$ , respectively). Error bars represent SEs.

**Figure 7: Effects of algorithm-induced biases on STRFs. (A)-(C)** Top row: NRC STRFs (Kn and Ks) for the same three units shown in Figures 5A-C (c.f. Fig.5A-C, bottom row). Middle row: NRC STRFs estimated from simulated responses to ml-noise (Knn) and song (Kns) when Kn is used as an LNP-type generative model for the neuron. Bottom row: NRC STRFs estimated from simulated responses to ml-noise (Ksn) and song (Kss) when Ks is used as an LNP-type generative model for the neuron. **(A)** An example cell for which NRC STRFs estimated from recorded responses to ml-noise (Kn) and song (Ks) and STRFs estimated from simulated responses to ml-noise and song stimuli (Knn, Kns, Ksn, and Kss) do not differ. **(B)** An example cell for which Kn and Ks differ significantly. These differences cannot be explained by biases

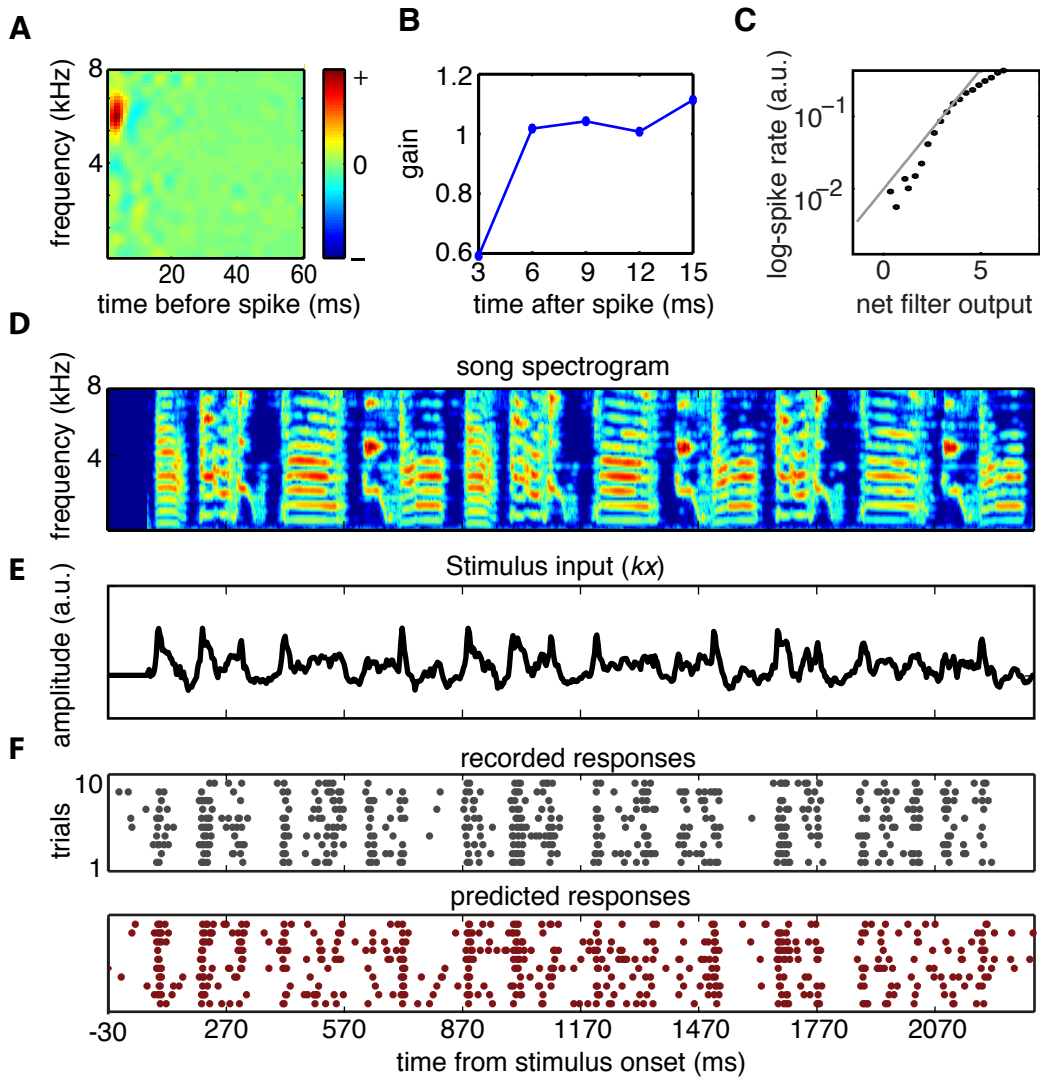
introduced by the estimation algorithm. **(C)** As in **(B)**, but in this case the differences between  $K_s$  and  $K_n$  can be explained by biases introduced by the estimation algorithm, since  $K_{ns}$  closely resembles  $K_s$  and not  $K_n$ . In addition,  $K_{sn}$  is more similar to  $K_s$  than  $K_n$ . **(D)-(F)** GLM STRFs for the same three units shown in panels A-C. **(G)** Additional example unit with broader spectral tuning and stronger inhibition. In all four examples the GLM reduces algorithm-induced biases (in all cases  $K_{ns}$  closely resembles  $K_n$  rather than  $K_s$  and  $K_{sn}$  is closer to  $K_s$  than to  $K_n$ ). **(H)** Distributions of similarity indices (SIs) between STRFs estimated from recorded and simulated data, for NRC (grey) and GLM (white). One indicates an exact match between STRFs and 0 indicates no correlation. Top: distribution of SIs computed between STRFs estimated from recorded responses to ml-noise ( $K_n$ ) and  $K_{ns}$  (STRFs estimated from simulated responses to song using  $K_n$  as a generative model for the neuron). Bottom: distributions of SIs computed between STRFs estimated from recorded responses to song ( $K_s$ ) and  $K_{sn}$  (STRFs estimated from simulated responses to ml noise stimuli using  $K_s$  as a generative model for the neuron). In both cases, the SI distributions for the GLM appear shifted to the right (and are centered closer to 1) when compared to the same distributions for NRC. In both cases, GLM and NRC distributions differ significantly ( $p < 0.001$ , two-sample KS test).

**Figure 8: Predictive power of the different models tested.** For both predictions of song and ml-noise responses, all the models that use a sparse prior or L1 regularizer (i.e., Exp GLM, Exp LN, spline GLM, and Linear L1) have an average prediction correlation that is significantly higher than the average prediction correlation for NRC, which uses a smoothing prior to regularize the STRFs. We found no significant difference in predictive power across the models that employ a sparse prior. Error bars show SEs.

Figure 1



# Figure 2





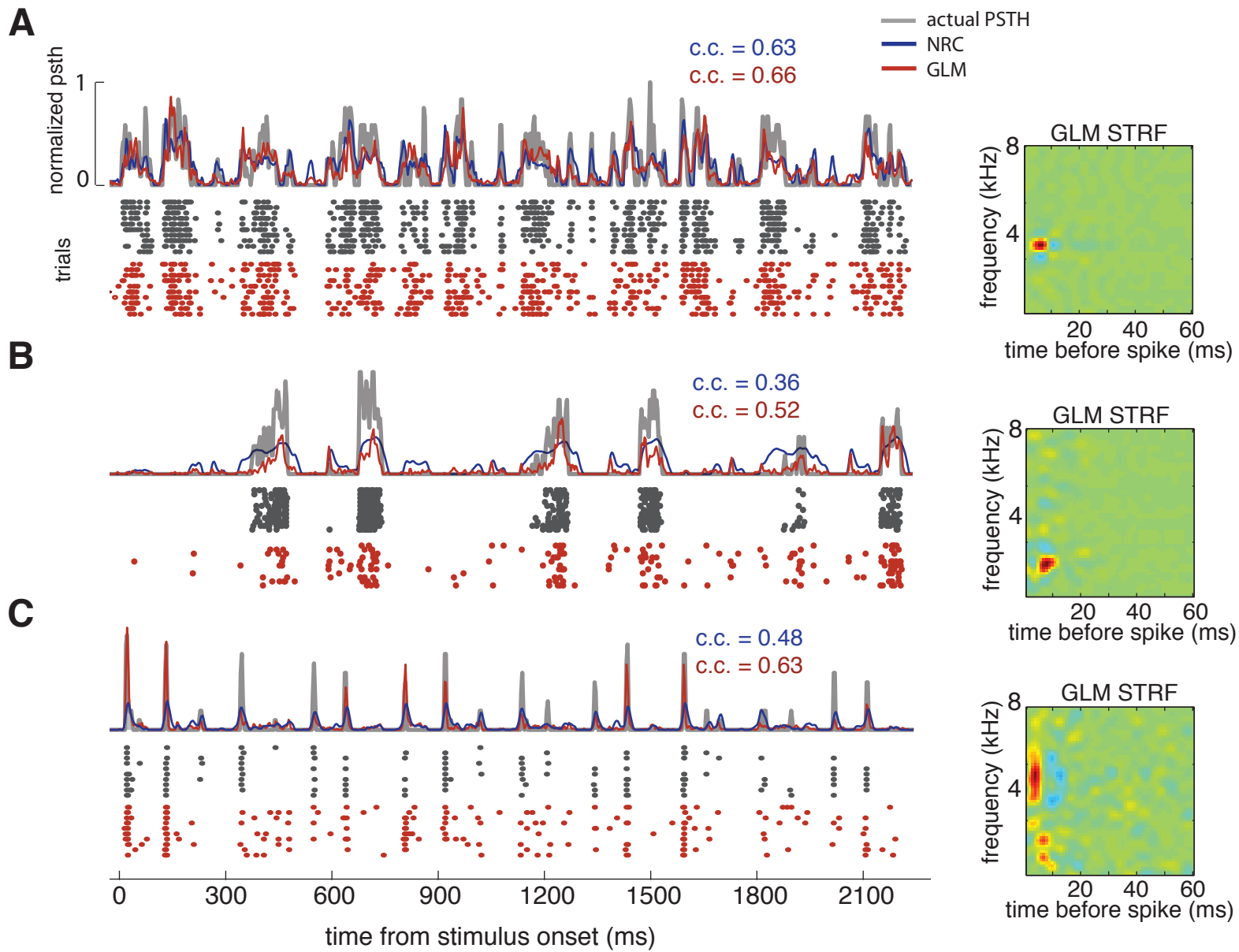
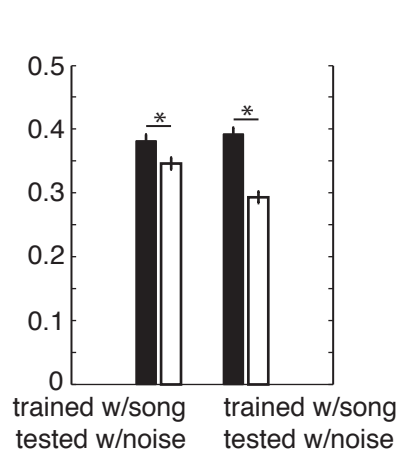
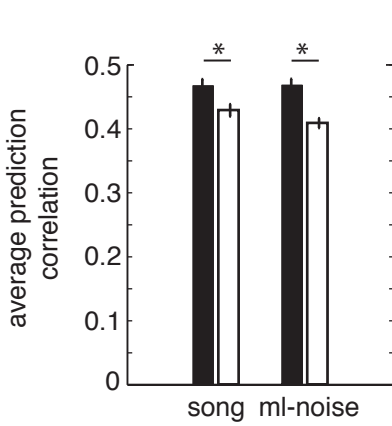
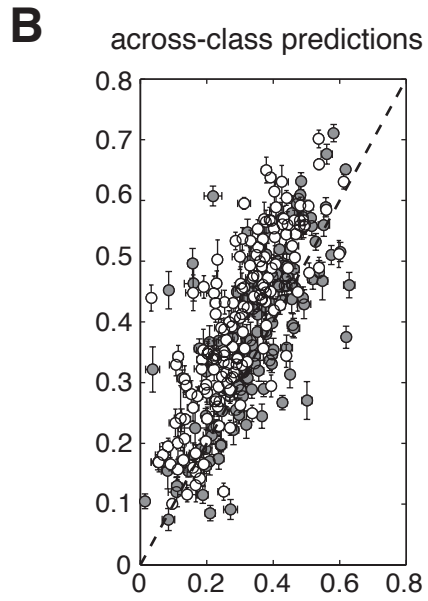
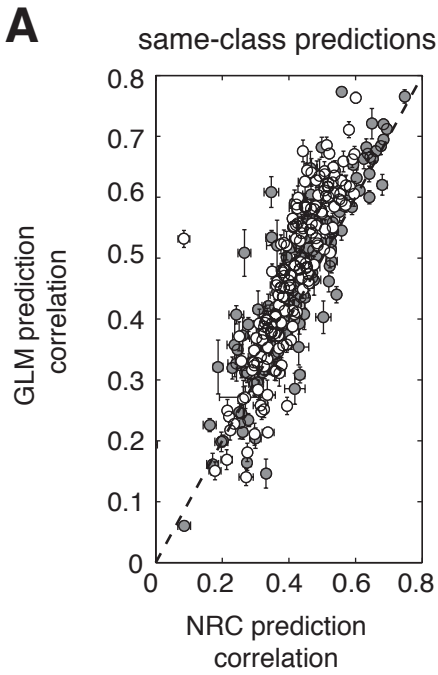
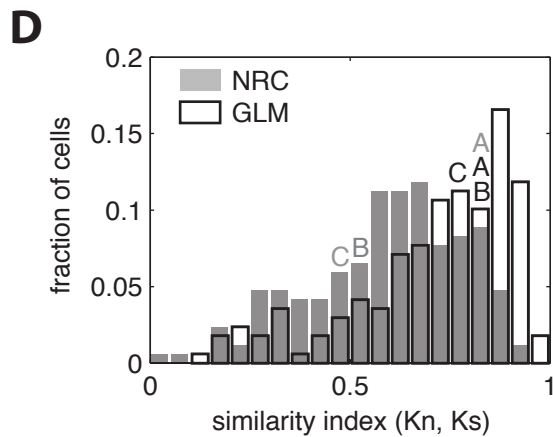
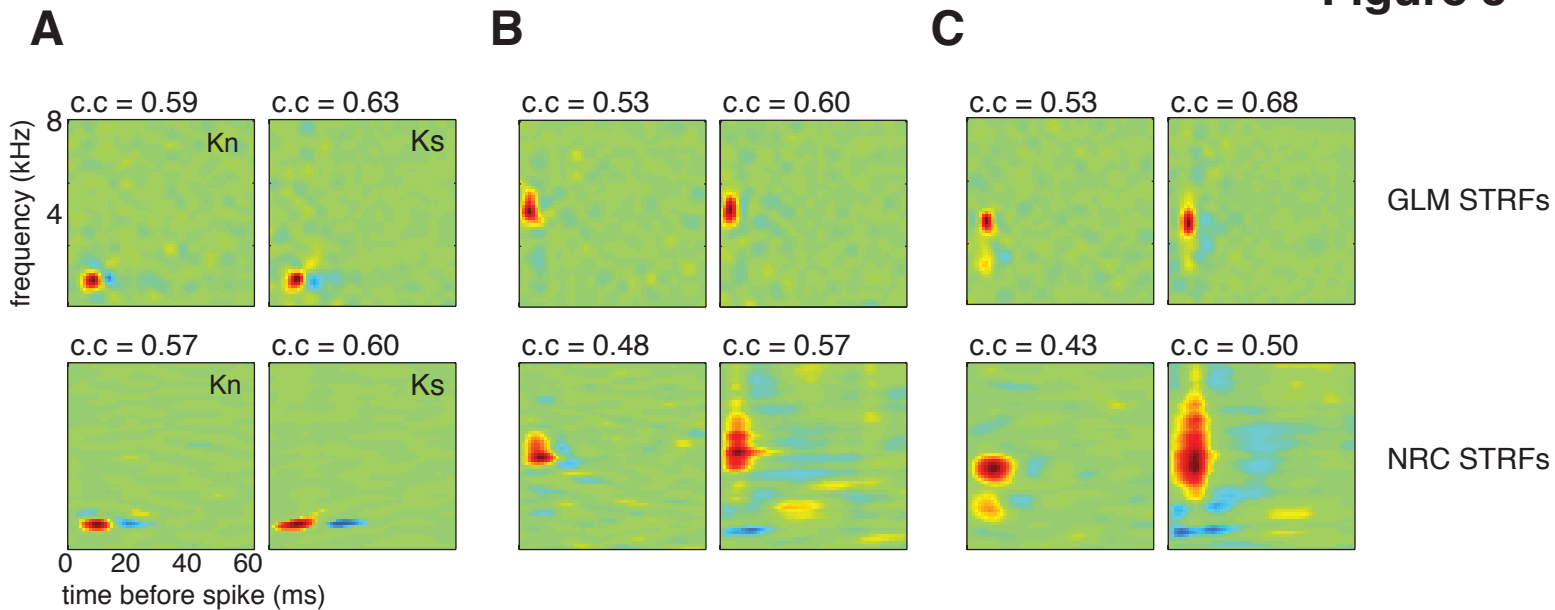
**Figure 3**

Figure 4



**Figure 5**



**Figure 6**

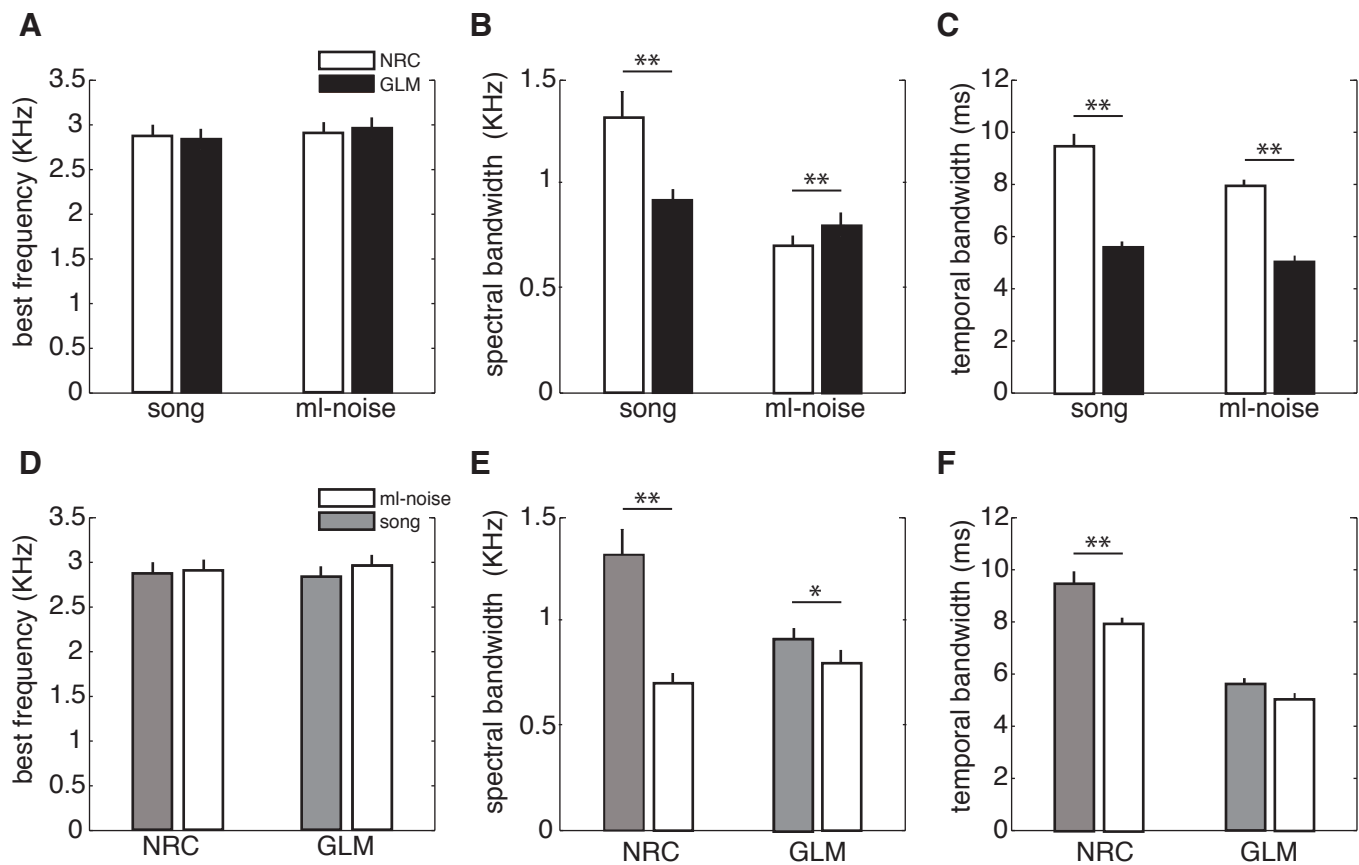


Figure 7

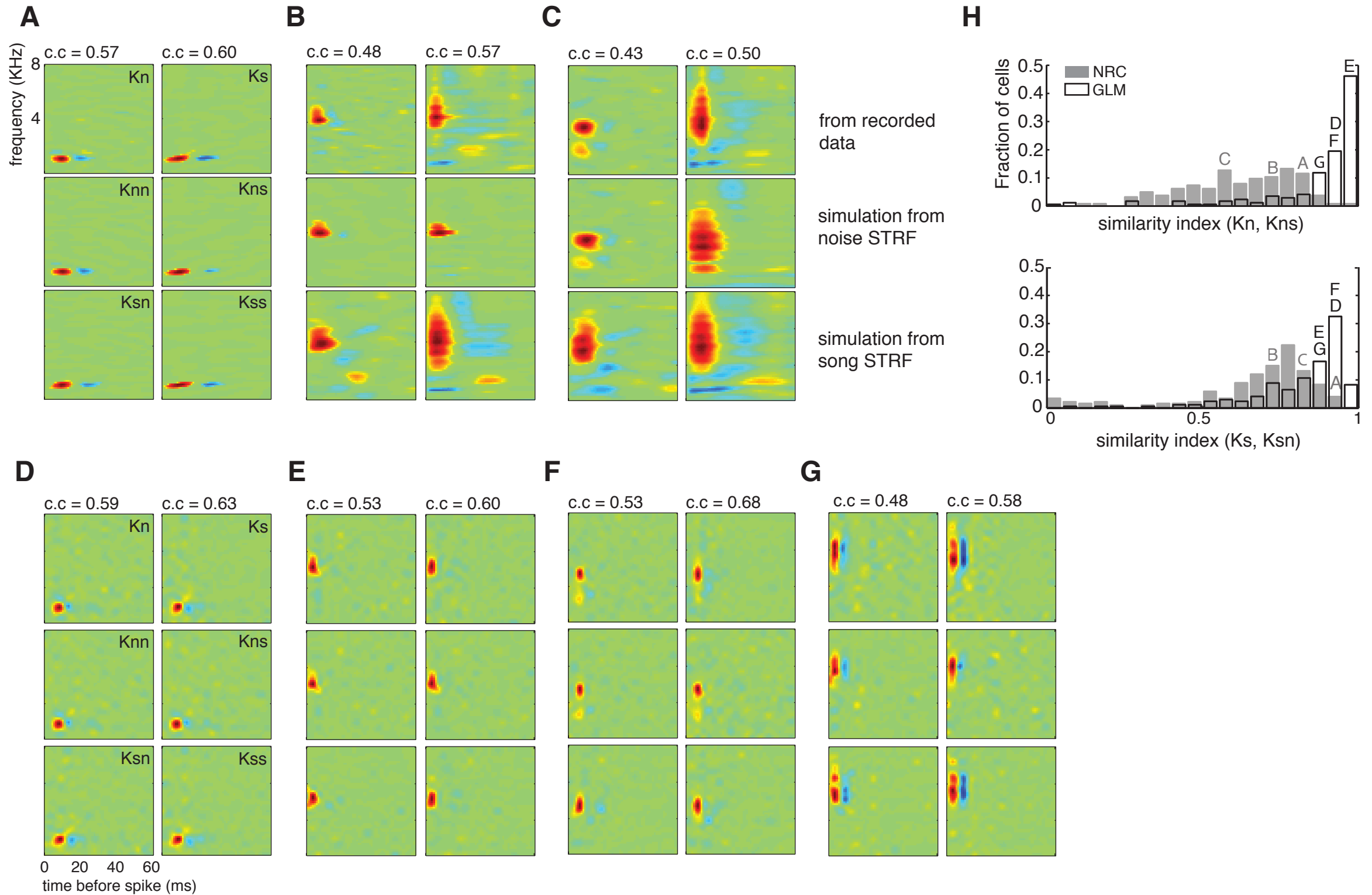


Figure 8

