Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods

Misha B. Ahrens[1], Jennifer F. Linden[2,3] and Maneesh Sahani[1]
[1]Gatsby Computational Neuroscience Unit, University College London
[2]Department of Anatomy and Developmental Biology, University College London
[3]Ear Institute, University College London

Manuscript submitted to *J. Neurosci.*

Abstract

The relationship between a sound and its neural representation in the auditory cortex remains elusive. Simple measures such as the frequency response area or frequency tuning curve provide little insight into the function of the auditory cortex in complex sound environments. Spectrotemporal receptive field (STRF) models, despite their descriptive potential, perform poorly when used to predict auditory cortical responses, showing that nonlinear features of cortical response functions, which are not captured by STRFs, are functionally important. We introduce a new approach to the description of auditory cortical responses, using multilinear modeling methods. These descriptions simultaneously account for several nonlinearities in the stimulus-response functions of auditory cortical neurons, including adaptation, spectral interactions, and nonlinear sensitivity to sound level. The models reveal multiple inseparabilities in cortical processing of time lag, frequency and sound level, and suggest functional mechanisms by which auditory cortical neurons are sensitive to stimulus context. By explicitly modeling these contextual influences, the models are able to predict auditory cortical responses more accurately than are STRF models. In addition, they can explain some forms of stimulus-dependence in STRFs that were previously poorly understood.

Introduction

The spectrotemporal receptive field (STRF, Aertsen et al., 1980) is widely used to characterise responses of primary auditory cortex (A1) neurons (e.g. deCharms et al., 1998; Depireux et al., 2001; Miller et al., 2002; Linden et al., 2003; Tomita and Eggermont, 2005). Recently, however, it has been shown that the STRF, when interpreted as a predictive model, is insufficient to account for auditory cortical responses (Sahani and Linden, 2003b; see also Machens et al., 2004), indicating that neurons in A1 respond nonlinearly to complex sounds.

Nonlinearities have profound consequences for the description of response properties. STRFs often change with the choice of stimulus (e.g. Theunissen et al., 2000; Blake and Merzenich, 2002; Valentine and Eggermont, 2004). While this might reflect adaptation to stimulus statistics, the apparent lability of STRF structure could also arise from the approximation of a nonlinear stimulus-response function with a linear function within different regions of stimulus space (Christianson et al., in press; Borst et al., 2005).

Thus, nonlinear models are required to gain a better understanding of auditory cortical responses to complex sounds. Here we introduce a new class of such models, which are compact and easy to estimate from limited experimental data. We show how these models can capture and quantify, in neuronal responses to complex stimuli, a number of nonlinear phenomena documented in experiments with isolated stimuli.

Nonlinear responses to sound level. The standard STRF model assumes that firing rate is related linearly to the sound level at every point in spectrotemporal space. This is not generally true (Phillips and Irvine, 1981). Our *input nonlinearity* model captures neuron-specific nonlinear mappings between spectrotemporal energy and neuronal responses to a complex sound.

Inseparabilities. Standard STRF models can describe time-frequency inseparabilities (i.e. interdependencies in temporal and spectral tuning properties), but not time-level or frequency-level inseparabilities, such as loudness-dependent latency (Phillips, 1989), or frequency-dependent tuning to loudness (Sutter, 2000). The input nonlinearity model allows us to estimate inseparabilities in *all* pairs of the three stimulus parameters, and thus to analyze how these inseparabilities affect neuronal responses to dynamic complex sounds.

Context effects. Suppressive phenomena such as two-tone and forward suppression are observed at all stages of the auditory pathway, including the auditory cortex (e.g. Brosch and Schreiner, 1997; Bartlett and Wang, 2005). Such phenomena are not modelled well with STRFs, and have therefore rarely been studied using complex sounds (but see Bar-Yosef et al., 2002). Extending the input nonlinearity model to the *context model*, we represent local contextual effects as the modulation of the effective sound level at a point in the spectrogram by the energy that falls within a short spectrographic window preceding that point. Thus, this spectrographic window is used to provide a local adaptive *context* for each element of the sonogram.

Each of these nonlinear models can be formulated and efficiently estimated using a common flexible framework based on multilinear analysis. These models improve on the predictive performance of STRF models by a factor of about 1.4. Moreover, for each neuron, the nonlinear models produce quantitative descriptions of the three phenomena described above. In addition, the context model sheds some light on the origin of the suppressive regions commonly observed in STRFs, and provides an explanation for certain systematic changes observed in STRFs measured under different stimulus conditions (Blake and Merzenich, 2002). Thus the models offer far more insight into auditory cortical response properties than has previously been available from STRF analysis.

Methods

The DRC stimulus.

The dynamic random chord stimuli used in these experiments have been described previously (Linden et al., 2003). Each stimulus was a series of 20 ms-long random chords — i.e., combinations of randomly selected tone pulses — played without intervening gaps. The centre frequencies of the tone pulses were chosen from 24 or 48 different possibilities (in the range 25-100 kHz or 2-32 kHz, respectively, depending on the approximate tuning of the neuron), spaced 1/12 octave apart. The number of tones constituting each chord was random, with an average of 2 tone pulses per octave per chord. Tone pulses were gated on and off with 5ms cosine ramps to reduce spectral splatter. The peak level of each pulse was chosen randomly from 10 different intensity levels, 5 dB-SPL apart in the range 25-70 dB-SPL. The frequency and intensity of each tone pulse were selected independently. Each of the two stimuli was 60s long, comprising 3000 different random chords. Exactly the same stimulus was repeated 10 (high-frequency) or 20 (low-frequency) times, with no intervening gap. All analyses are based on the final 9 or 19 repetitions, so as to avoid artifacts due to strong adaptation at the onset of stimulation.

Experimental methods.

The experimental methods were similar to those described by Linden et al. (2003). Surgical procedures conformed to protocols approved by the University of California at San Francisco's Committee on Animal Research and were in accordance with US federal guidelines for care and use of animals in research. Mice and rats were anesthetized and maintained at a surgical plane of anesthesia with ketamine and medetomidine or pentobarbital, and extracellular recordings were made in early auditory cortical areas (primary auditory cortex, A1, of rats, and A1 and anterior auditory field, AAF, of mice). Recordings targeted thalamorecipient layers III/IV (Smith and Populin, 2001) by cortical depth (350-600 μm below the dural surface) and by the polarity and size of stimulus-evoked local field potentials. The recordings were later band-pass filtered and then analyzed using Bayesian spike-sorting techniques (Lewicki, 1994) (user interface software by M. Kvale, UCSF) to extract responses from single units or small clusters of neurons.

Notation

In this paper, bold-faced san-serif letters such as $\mathbf{w}$ represent arrays: these may be vectors, matrices, or higher-order cartesian tensors. Bold-faced superscripts are used with arrays of weights and stimuli to indicate the corresponding dimensions. Thus $\mathbf{w}^{\mathbf{tf}}$ represents an array of STRF weights that span time-lag and frequency. Elements in the array are italicised and subscripted. Thus $w_{jk}^{\mathbf{tf}}$ is the element at time lag j and frequency bin k in the matrix $\mathbf{w}^{\mathbf{tf}}$. To simplify notation, we have adopted modified versions of operators from multilinear algebra. The symbol $\otimes$ generalises the vector outer product, such that, for example, if $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ are all vectors then $\mathbf{a} = \mathbf{b} \otimes \mathbf{c} \otimes \mathbf{d}$ is a three-dimensional array, with elements $a_{ijk} = b_i c_j d_k$. The symbol $\bullet$ generalises the inner product, so that all indices shared between arguments on the left and right sides of the operator are summed (or *contracted*); which indices are shared will be clear from the context. Standard matrix multiplication is shown without an explicit operator.

Model estimation

To capture the response properties of auditory cortical neurons, predictive models may be used to approximate the neural stimulus-response function, i.e., the transformation from stimulus to neuronal firing rates. A widely used model is based on the spectrotemporal receptive field (STRF), here denoted $\mathbf{w}^{\mathbf{tf}}$, where the superscripts

denote time-lag and frequency respectively. Such models predict the time-varying firing rate of a neuron, $r(i)$ (where i indexes time bins), from the stimulus $s(i,k)$ (denoting the power in time bin i and frequency bin k of the spectrographic representation of the sound) through the formula

$$\hat{r}(i) = c + \sum_{j=1}^{J}\sum_{k=1}^{K} w_{jk}^{\mathbf{tf}} s(i-j+1,k) \tag{1}$$

The parameters of the model — the constant c and the elements of the STRF matrix $\mathbf{w^{tf}}$ — should be chosen so as to make the predicted firing rate $\hat{r}(i)$ as close as possible to the observed firing rate $r(i)$. One approach is to present the same stimulus s multiple times; the mean firing rate over repeated trials will then be approximately Gaussian distributed around the "true" mean stimulus-dependent firing rate. The parameters c and $\mathbf{w^{tf}}$ can then be found by least-squares linear regression, minimising the squared error $\mathcal{E} = \sum_i (r(i) - \hat{r}(i))^2$ (e.g. Machens et al., 2004), which is equivalent to maximum-likelihood estimation under the Gaussian assumption.

One danger of maximum likelihood regression is overfitting. That is, the parameters which minimise the squared error on one set of *training* data, may do so by capturing stimulus-unrelated variability in those measurements, and therefore perform poorly when predicting responses to a novel set of *test* data. If this happens, the features exhibited by the best-fit model cannot be presumed to reflect actual structure in the neural response function.

Overfitting may be contained by regularisation methods, in which extra terms are introduced to the objective function (or equivalently, a prior belief about the parameter properties is expressed) so as to penalize solutions that have, for example, many large parameter values, or large variation in parameter values. These extra terms, for example describing the scale of the smoothness, are often set by hand. Here we used a data-driven method called Automatic Smoothness Determination (ASD) to estimate optimal values for the scale of the smoothness and the magnitude of the model parameters (Sahani and Linden, 2003a).

Bilinear models

Separable STRF models. A linear spectrotemporal receptive field $\mathbf{w^{tf}}$ is called separable if, in its matrix form, it can be written as the outer product of two vectors, $\mathbf{w^{tf}} = \mathbf{w^t}\mathbf{w^f}^{\mathsf{T}}$; or, in terms of the components, $w_{jk}^{\mathbf{tf}} = w_j^{\mathbf{t}} w_k^{\mathbf{f}}$. Functionally, this implies that the relative response of the neuron to tones of different frequencies, given by a vector $\mathbf{w^f}$, is preserved across time-lag, with the absolute responses scaled by a different (possibly negative) factor at each time, given by the vector $\mathbf{w^t}$; or equivalently, that the temporal response profile at all frequencies is identical up to a per-frequency scale factor. Mathematically, the predicted firing rate in time bin i is, up to an additive constant (discussed later),

$$\widehat{r}(i) = \sum_{j=1}^{J}\sum_{k=1}^{K} w_j^{\mathbf{t}} w_k^{\mathbf{f}} s(i-j+1,k).$$

Thus, although the separable model is nonlinear in the parameters taken all together, it is linear in each of the vectors $\mathbf{w^t}$ and $\mathbf{w^f}$ alone. Such a model is called *bilinear*. A graphical illustration of a bilinear model is shown in Fig. 1A.

Many STRFs in the auditory system appear to be well-modelled as separable (Linden et al., 2003; Simon et al., 2007), but even when this is not the case, a bilinear model may be used to *approximate* the full inseparable linear (in $\mathbf{w^{tf}}$) relationship. Here, we discuss this bilinear approximation to the STRF as a preliminary to the more extensive multilinear models introduced later.

One way to fit a bilinear model is to estimate the full receptive field $\mathbf{w^{tf}}$, and then examine the singular value decomposition (SVD; Strang, 1988) of the resulting matrix. The component vectors corresponding to the

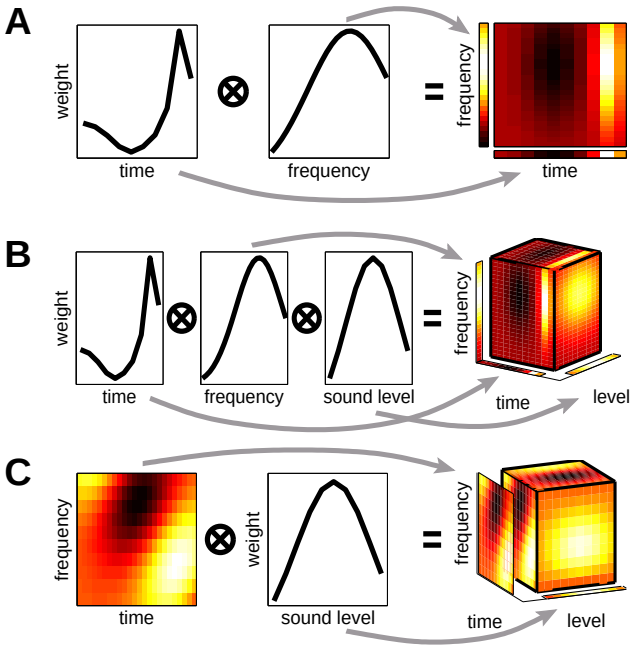

Figure 1: ***A:*** A bilinear STRF. The STRF (right) is equal to the outer product of two vectors, one describing the temporal response (left) and the other describing the spectral response (middle). ***B:*** Schematic of the fully separated input nonlinearity model $\mathbf{w^t} \otimes \mathbf{w^f} \otimes \mathbf{w^l}$. The response to latency, frequency and sound level is assumed to be fully separable. ***C:*** Schematic of the input nonlinearity model $\mathbf{w^{tf}} \otimes \mathbf{w^l}$. The full spectro-temporal component (left) and the sound-level component (middle) produce a response function (right) that is separable in sound-level and time, sound-level and frequency, but not in time and frequency. Note that there are weights (not shown) in the interior of the cube.

largest singular value give the best separable approximation to the full STRF matrix in the least-squares sense; that is, they minimise the term $\sum_{jk} |w^{\mathbf{tf}}_{jk} - w^{\mathbf{t}}_j w^{\mathbf{f}}_k|^2$. However, even when $\mathbf{w^{tf}}$ is itself found by least-squares regression, this cost function is different from the bilinear model error

$$\mathcal{E} = \sum_i (r(i) - \hat{r}(i))^2 = \sum_i \left(r(i) - \sum_{jk} w^{\mathbf{t}}_j w^{\mathbf{f}}_k s(i-j+1,k) \right)^2 .$$

We derive algorithms to minimise this model error (possibly with additional regularisation terms) directly.

It is helpful to simplify notation as follows. Consider an expanded three-dimensional stimulus array, augmented by the addition of a time-lag dimension: $M^{\mathbf{itf}}_{ijk} = s(i-j+1,k)$. This change of notation allows us to write the bilinear spike rate prediction as

$$\widehat{r}(i) = \sum_{jk} w^{\mathbf{t}}_j w^{\mathbf{f}}_k M^{\mathbf{itf}}_{ijk}, \quad \text{or} \quad \widehat{\mathbf{r}} = (\mathbf{w^t} \otimes \mathbf{w^f}) \bullet \mathbf{M^{itf}}, \tag{2}$$

where the second form adapts the tensor notation of multilinear algebra as described above. Here, since $\mathbf{M^{itf}}$ shares time-lag and frequency indices (j and k) with $(\mathbf{w^t} \otimes \mathbf{w^f})$, the sum implied by the $\bullet$ operator is over those two dimensions, with the third index i remaining, corresponding to the bin-by-bin predictions collected in the vector $\widehat{\mathbf{r}}$. If we now write the vector norm as $\|\mathbf{a}\| = \sqrt{\mathbf{a} \bullet \mathbf{a}} = \left(\sum_i |a_i|^2\right)^{1/2}$, the squared-error cost function becomes

$$\mathcal{E} = \left\| \mathbf{r} - (\mathbf{w^t} \otimes \mathbf{w^f}) \bullet \mathbf{M^{itf}} \right\|^2 .$$

In the following we will use component notation, tensor notation or both as needed to aid clarity.

Parameter estimation. At a local minimum of the error, we have

$$\frac{\partial}{\partial w_j^{\mathbf{t}}}\mathcal{E} = 0, \quad \text{and} \qquad \frac{\partial}{\partial w_k^{\mathbf{f}}}\mathcal{E} = 0.$$

Differentiating and rearranging yields the fixed point conditions

$$\sum_{ik} M_{ijk}^{\mathbf{itf}} w_k^{\mathbf{f}} r(i) = \sum_{ij'k'k} M_{ij'k'}^{\mathbf{itf}} w_{j'}^{\mathbf{t}} w_{k'}^{\mathbf{f}} M_{ijk}^{\mathbf{itf}} w_k^{\mathbf{f}}, \qquad \text{or} \quad (\mathbf{w^f} \otimes \mathbf{r}) \bullet \mathbf{M^{itf}} = \big((\mathbf{w^f} \otimes \mathbf{w^t}) \bullet \mathbf{M^{itf}}\big) \bullet \big(\mathbf{w^f} \bullet \mathbf{M^{itf}}\big),$$

and

$$\sum_{ij} M_{ijk}^{\mathbf{itf}} w_j^{\mathbf{t}} r(i) = \sum_{ij'jk'} M_{ij'k'}^{\mathbf{itf}} w_{j'}^{\mathbf{t}} w_{k'}^{\mathbf{f}} M_{ijk}^{\mathbf{itf}} w_j^{\mathbf{t}}, \qquad \text{or} \quad (\mathbf{w^t} \otimes \mathbf{r}) \bullet \mathbf{M^{itf}} = \big((\mathbf{w^f} \otimes \mathbf{w^t}) \bullet \mathbf{M^{itf}}\big) \bullet \big(\mathbf{w^t} \bullet \mathbf{M^{itf}}\big).$$

Defining $\mathbf{A} = \mathbf{w^f} \bullet \mathbf{M^{itf}}$ and $\mathbf{B} = \mathbf{w^t} \bullet \mathbf{M^{itf}}$, and noting that both $\mathbf{A}$ and $\mathbf{B}$ are matrices, we obtain the matrix equations

$$\mathbf{A}^\mathsf{T}\mathbf{r} = \mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{w^t} \quad \text{and} \quad \mathbf{B}^\mathsf{T}\mathbf{r} = \mathbf{B}^\mathsf{T}\mathbf{B}\mathbf{w^f},$$

which can be solved to give

$$\mathbf{w^t} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{r} \quad \text{and} \quad \mathbf{w^f} = (\mathbf{B}^\mathsf{T}\mathbf{B})^{-1}\mathbf{B}^\mathsf{T}\mathbf{r}. \tag{3}$$

It should come as no surprise that these each resemble the solution to a linear regression problem, as the bilinear model is linear in each parameter vector separately. However, the matrix $\mathbf{A}$ ($\mathbf{B}$) is itself a function of $\mathbf{w^f}$ ($\mathbf{w^t}$), and so these equations do not give a closed-form solution to the optimisation problem. Instead, the equations are applied iteratively, alternating between updates for $\mathbf{w^t}$ and for $\mathbf{w^f}$. Each iteration decreases the squared error and, since $\mathcal{E}$ cannot drop below zero, the iterations must converge to an optimum in the parameter space.

The background rate. Since neurons often fire in the absence of a stimulus, a complete predictive model must include a constant term. As in the linear case, this can be introduced cleanly into the multilinear framework by augmenting the stimulus appropriately. If there are T time points in the stimulus, and the time-lag index ranges from $1 \ldots J$ and the frequency index from $1 \ldots K$, so that $\mathbf{M^{itf}}$ is $T \times J \times K$, we consider a $T \times (J+1) \times (K+1)$ array $\mathbf{Q^{itf}}$ with

$$Q_{ijk}^{\mathbf{itf}} = \begin{cases} M_{ijk}^{\mathbf{itf}} & j \le J \quad k \le K \\ 1 & j = J+1 \quad k = K+1 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then, with $\mathbf{w^t}$ and $\mathbf{w^f}$ also augumented to contain $J+1$ and $K+1$ elements respectively, the model $\mathbf{r} = (\mathbf{w^t} \otimes \mathbf{w^f}) \bullet \mathbf{Q^{itf}}$ becomes equivalent to a model with background rate, $\mathbf{r} = c + (\mathbf{w^t} \otimes \mathbf{w^f}) \bullet \mathbf{M^{itf}}$ with $c = w_{J+1}^{\mathbf{t}} w_{K+1}^{\mathbf{f}}$.

Multilinear models

We have seen that a separable STRF may be viewed as a bilinear predictive model. Shortly, we will see that more complex models, incorporating nonlinear level sensitivities and acoustic-context-dependent sensitivities, can be also be expressed in a multilinear form. However, before going on to develop these models, it is useful to discuss some general aspects of parameter estimation in the multilinear setting (see also Ahrens et al., in press).

The general predictive form of a multilinear model can be written

$$\widehat{r}(i) = \sum_{jk\ldots m} a_j b_k \ldots z_m Q_{ijk\ldots m}, \quad \text{or} \quad \widehat{\mathbf{r}} = (\mathbf{a} \otimes \mathbf{b} \otimes \ldots \otimes \mathbf{z}) \bullet \mathbf{Q}, \tag{5}$$

where $\mathbf{a}$, $\mathbf{b}$, ..., $\mathbf{z}$ are parameter vectors and $\mathbf{Q}$ is a fixed multidimensional array that depends only on the stimulus.

Alternating least squares. As in the case of the bilinear model discussed above, the squared error

$$\mathcal{E} = \|\mathbf{r} - (\mathbf{a} \otimes \mathbf{b} \otimes \ldots \otimes \mathbf{z}) \bullet \mathbf{Q}\|^2 \tag{6}$$

can be minimised by cycling through a set of update equations, each of which resembles the solution to a linear regression problem:

$$\begin{aligned} \mathbf{A} &= (\mathbf{b} \otimes \mathbf{c} \otimes \ldots \otimes \mathbf{z}) \bullet \mathbf{Q} & \mathbf{a} &= (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{r} \\ \mathbf{B} &= (\mathbf{a} \otimes \mathbf{c} \otimes \ldots \otimes \mathbf{z}) \bullet \mathbf{Q} & \mathbf{b} &= (\mathbf{B}^\mathsf{T}\mathbf{B})^{-1}\mathbf{B}^\mathsf{T}\mathbf{r} \\ & \vdots & & \\ \mathbf{Z} &= (\mathbf{a} \otimes \mathbf{b} \otimes \ldots \otimes \mathbf{y}) \bullet \mathbf{Q} & \mathbf{z} &= (\mathbf{Z}^\mathsf{T}\mathbf{Z})^{-1}\mathbf{Z}^\mathsf{T}\mathbf{r} \end{aligned} \tag{7}$$

Each of these updates minimises the squared error with respect to one parameter vector, whilst holding the others fixed. Thus, the algorithm, a variant of *alternating least squares*, is guaranteed not to increase the squared error at any step, and hence (as the error is bounded below by zero) to converge.

Control of overfitting. As the number and length of the parameter vectors in any model increases, so does the danger of overfitting. As described above, *regularisation* is the practice of adding terms to the cost function to be optimised, so as to discourage overfitting. Here, we adopt a Bayesian perspective, in which the simple cost function corresponds to the (log) likelihood of the parameters given the data, and the regularisation terms express our prior beliefs about their forms or values.

The objective function used is of the form

$$\mathcal{E}_r = \mathcal{E} + \mathbf{a}^\mathsf{T}\mathbf{D}_\mathbf{a}\mathbf{a} + \mathbf{b}^\mathsf{T}\mathbf{D}_\mathbf{b}\mathbf{b} + \ldots$$

Viewed probabilistically, the squared-error term $\mathcal{E}$ corresponds to a Gaussian likelihood. Although a likelihood based more directly on a mean of Poisson terms may be a better model for empirical firing rate data, the Gaussian assumption makes fitting algorithms more tractable and is appropriate in the regime of many observations or moderate to high firing rates, where the trial-averaged Poisson-based likelihood approaches the Gaussian form. The additional terms of the objective function express zero-mean Gaussian priors on each parameter vector, with covariances $\mathbf{D}_\mathbf{a}^{-1}$ etc. The priors are zero-mean because we have no reason to expect *a priori* parameters of any particular sign. The covariance matrices are chosen to favor smaller parameter values (which we refer to as "shrinkage") as well as smoothness in the parameter vectors.

The alternating least squares algorithm of (7) can be adapted to minimise $\mathcal{E}_r$ through a small modification. Given a matrix $\mathbf{A}$ as defined in (7), the update for $\mathbf{a}$ becomes

$$\mathbf{a} = (\mathbf{A}^\mathsf{T}\mathbf{A} + \mathbf{D}_\mathbf{a})^{-1}\mathbf{A}^\mathsf{T}\mathbf{r},$$

and similarly for the other parameter vectors. For fixed prior covariances, these updates are guaranteed to converge, as before.

To obtain appropriate prior covariance matrices $\mathbf{D}_{(\cdot)}$, we adapted the ASD scheme of Sahani and Linden (2003a). In this scheme, each prior covariance matrix is parametrised by two parameters $\theta_{1,2}^{(\cdot)}$, describing the degree of shrinkage and smoothing; optimal values of the parameters are found through an automatic procedure. Here, we used this procedure within the first few iterations of the alternating least squares framework. If ASD updates are used at every step, the objective function varies between iterations (because the $\mathbf{D}$ matrices change), and convergence of the fitting procedure is no longer guaranteed. For this reason, the covariance parameters were only updated during the first three iterations. Once these covariance parameters were fixed, the remaining iterations were again guaranteed to converge.

Estimation error. Error bars for the parameters were obtained by a bootstrap procedure (Effron and Tibshirani, 1993). Parameters were re-fit ten times, in each case using a equinumerous training set randomly redrawn, with replacement, from the available data. The error bars in all figures indicate the standard deviations of these ten estimates.

The input nonlinearity model

In this and the next sections we develop predictive auditory models that express various non-linear stimulus dependencies in multilinear form, allowing them to be fit using the methods described above.

Stimulus representation. An auditory STRF model is linear in the sonogram. Consequently, the predictions of the model, and the parameter values obtained, depend on whether the input stimulus $s(i,k)$ represents the intensity, the power or the log-intensity of the stimulus at the given time and frequency. Indeed, the best predictions may be obtained not from any of these external representations, but rather from something closer to the cell's rate-level function, or some other cell-specific mapping. Here, we consider a generalised model in which a suitable non-linear transform of the stimulus can be found directly from the data. Based on the STRF model of equation 1, such a model has the form

$$\widehat{r}(i) = c + \sum_{jk} w^{\mathbf{tf}}_{jk} g(s(i-j+1,k)), \tag{8}$$

where g is a function to be learnt. The mapping g is an "input nonlinearity", which transforms the representation of sound level in the spectrogram before it is spectro-temporally filtered by $\mathbf{w}^{\mathbf{tf}}$. This form is linear in $\mathbf{w}^{\mathbf{tf}}$; but g has not yet been parameterised so as to make estimation possible. One natural parameterisation, common in the regression literature, is as a linear combination of a fixed set of basis functions $\{g_l\}$, so that $g(x) = \sum_l w^{\mathbf{l}}_l g_l(x)$, for some parameter vector $\mathbf{w}^{\mathbf{l}}$.

For the stimulus considered here, the level of each tone pulse was drawn from a set of 10 distinct possible values. Thus, a natural set of basis functions comprises 10 indicator functions, each taking the value 1 for a single possible input sound level, and 0 otherwise. This leads to a simple interpretation of $w^{\mathbf{l}}_l$ as the net effective input correponding to a pulse at the l^{th} intensity level. However, the mathematical development that follows is applicable to any choice of basis, and thus similar methods could be used with other stimuli, even if not discretised in level (Ahrens et al., in press).

Given this basis parameterisation, (8) can be re-written

$$\widehat{r}(i) = c + \sum_{jkl} w^{\mathbf{tf}}_{jk} w^{\mathbf{l}}_{l}\, g_l(s(i-j+1,k)). \tag{9}$$

If we now define a four-index array $M^{\mathbf{itfl}}_{ijkl} = g_l(s(i-j+1,k))$, and consider a separable STRF model $w^{\mathbf{tf}}_{jk} = w^{\mathbf{t}}_{j} w^{\mathbf{f}}_{k}$, the model can be written in multlinear form

$$\widehat{r}(i) = c + \sum_{jkl} w^{\mathbf{t}}_{j} w^{\mathbf{f}}_{k} w^{\mathbf{l}}_{l}\, M^{\mathbf{itfl}}_{ijkl} \quad \text{or} \quad \widehat{\mathbf{r}} = (\mathbf{w^t} \otimes \mathbf{w^f} \otimes \mathbf{w^l}) \bullet \mathbf{Q^{itfl}}, \tag{10}$$

with a four-dimensional array $\mathbf{Q^{itfl}}$ defined by augmenting $\mathbf{M^{itfl}}$ in a manner analogous to (4). This parameter grouping is illustrated in Fig. 1B.

Grouping stimulus dimensions. The unseparated form (9) is itself a bilinear model, with linear dependence on each of the STRF parameters $w^{\mathbf{tf}}_{jk}$; that is, it can be written

$$\widehat{\mathbf{r}} = (\mathbf{w^{tf}} \otimes \mathbf{w^l}) \bullet \mathbf{Q^{itfl}}$$

as illustrated in Fig. 1C. It is therefore possible to apply the multilinear estimation framework directly, without forcing the time-frequency component of the model to be separable. Formally, this may be done by considering the "rasterised" or "unrolled" vector form of the matrix $\mathbf{w^{tf}}$, often written $\mathsf{vec}\left(\mathbf{w^{tf}}\right)$, while similarly unrolling the corresponding dimensions of the array $\mathbf{Q^{itfl}}$.

This model has many more parameters than the fully separated, trilinear version (10), and is thus considerably more prone to overfitting, making regularisation yet more crucial. In this case, the automatic smoothness determination procedure (Sahani and Linden, 2003b) obtains separate smoothing parameters for the time-lag and frequency dimensions, even though they are combined into a single vector.

Although this model corresponds well to the standard STRF representation, it is clearly not the only way in which parameters of the trilinear model may be grouped. Instead, a neuron might be better described as having an inseparable frequency-level receptive field, described by a matrix $\mathbf{w^{fl}}$. In this case the model is

$$\widehat{\mathbf{r}} = (\mathbf{w^t} \otimes \mathbf{w^{fl}}) \bullet \mathbf{Q^{itfl}} \tag{11}$$

This model is able to capture inseparable frequency-level components of the response function, as are often seen in static receptive fields, but are ignored in linear STRF models. However, the model assumes that the time component of the response function is separable from the frequency-level component.

A neuron might also have an inseparable time-sound level component, in which case the appropriate model might be

$$\widehat{\mathbf{r}} = (\mathbf{w^{tl}} \otimes \mathbf{w^f}) \bullet \mathbf{Q^{itfl}} \tag{12}$$

which is able to capture inseparable time-level properties of the response function, i.e. changes in sound level tuning at different lag times.

For any given neuron, we can consider each of these possible models, evaluating the performance of each by cross-validation and choosing the most predictively successful model to characterize the cell.

Indeed, it is possible, in principle, to dispense with separability all together, fitting a linear model with a three-dimensional parameter array $\mathbf{w^{tfl}}$. In practice, however, the number of parameters implied by this model (the product of the numbers of time-lags, of frequency bins and of level basis functions) makes it impractical to estimate without considerable overfitting.

Adaptation and two tone interaction: the context model

Documented nonlinear effects in the auditory pathway include two-tone suppression, where the response evoked by a tone is affected by a simultaneously presented tone at a different frequency (e.g. Sutter and Schreiner, 1991); and forward suppression, where the tone-evoked response is modulated by earlier tones at the same frequency (e.g. Brosch and Schreiner, 1997).

These and other *contextual* effects are partly modelled by the standard STRF; however this framework forces the contextual influence to be additive. Here, we consider multilinear models that also describe multiplicative contextual effects. In particular, we consider models in which the context of a tone pulse in the DRC — or, more generally, of the energy at a point in the spectrogram, which we call a *time-frequency element* — may be viewed as multiplicatively modulating its effective "level"; that is, the stimulus in time bin i at frequency k has an effective strength given by

$$g(s(i,k)) \cdot (c_2 + \mathsf{Context}(i,k)). \tag{13}$$

Here, g is a nonlinear function of the type considered earlier (equation 8). $\mathsf{Context}(i,k)$ is a function mapping the stimulus context around the time-frequency point under consideration to a multiplicative factor.

We assume that the contextual modulator $\mathsf{Context}(i,k)$ is itself a multilinear function of the same form as the input nonlinearity model, but indexed relative to the time and frequency being modulated, as sketched in Fig. 2. That is, we write the contextual modulation as

$$\mathsf{Context}(i,k) = \sum_{\substack{m=1 \\ (m,n)\neq(1,\Phi)}}^{M} \sum_{n=1}^{N} \sum_{p=1}^{P} w_m^{\boldsymbol{\tau}} w_n^{\boldsymbol{\phi}} w_p^{\boldsymbol{\lambda}} h_p(s(i-m+1, k-\Phi-1+n)),$$

where $\Phi = (N-1)/2$ is the maximal absolute frequency difference considered between the contextual and modulated time-frequency elements, and the exclusion of $(m,n) = (1,\Phi)$ is due to the fact that a time-frequency element cannot be in its own context (that is, $s(i,k)$ is not part of the expression for $\mathsf{Context}(i,k)$). The vectors $\mathbf{w}^{\boldsymbol{\tau}}$ and $\mathbf{w}^{\boldsymbol{\phi}}$ contain weights that depend on the relative time differences $m = 1 \ldots M$ and frequency differences $n = 1 \ldots N$ respectively. The vector $\mathbf{w}^{\boldsymbol{\lambda}}$ transforms the contextual sound energy in terms of a set of P basis functions $h_p(s)$, which are here chosen to be identical to the $g_l(s)$ used in the input nonlinearity model, although they may differ in general.

In the tensor notation used before,

$$\mathsf{Context}(i,k) = (\mathbf{w}^{\boldsymbol{\tau}} \otimes \mathbf{w}^{\boldsymbol{\phi}} \otimes \mathbf{w}^{\boldsymbol{\lambda}}) \bullet \mathbf{M}^{\boldsymbol{\tau\phi\lambda}}(i,k), \tag{14}$$

where $\mathbf{M}^{\boldsymbol{\tau\phi\lambda}}(i,k)$ is a stimulus array which depends on the (i,k) position of the time-frequency element being modulated, and (if the bases $\{g_l\}$ and $\{h_p\}$ are the same)

$$[M^{\boldsymbol{\tau\phi\lambda}}(i,k)]_{mnp} = \begin{cases} 0 & \text{if } (m,n) = (1,\Phi) \\ M^{\mathbf{itfl}}_{im(k-\Phi-1+n)p} & \text{otherwise} \end{cases} \tag{15}$$

for $1 \le m \le M$; $1 \le n \le N$; $1 \le p \le L$; with $\mathbf{M}^{\mathbf{itfl}}$ defined as for (10). $\mathbf{M}^{\boldsymbol{\tau\phi\lambda}}(i,k)$ will be referred to the "contextual subunit" of the time-frequency element at time i and frequency k (Fig. 2).

The contextual influence modulates the effectiveness of the stimulus $s(i,k)$ according to (13). The modulated sound strengths are then combined according to the non-contextual multilinear form. Thus the full model

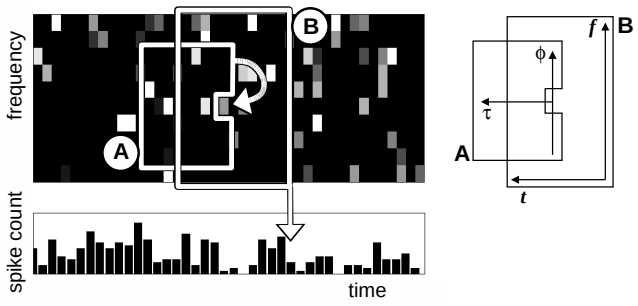

Figure 2: Schematic of the context model for a DRC stimulus. The sound level components are not shown. ***A:*** The context part of the model, dependent on time-frequency (i, f) and parametrised by (τ, ϕ, λ). A weighted sum of the tone pulses (or, for a general sound, time-frequency elements) in the box acts to modulate the effective sound level of the target tone, indicated by the arrow. ***B:*** the time-frequency part of the model, equivalent to the domain of action of the STRF. A weighted sum of the amplitudes of the tone pulses in the box after modulation by context (A) predicts the spike rate, indicated by the downward pointing arrow. ***Inset:*** The axes labels for A and B.

becomes:

$$\widehat{r}(i) = c + \sum_{jkl} \underbrace{w^{\mathbf{t}}_j w^{\mathbf{f}}_k w^{\mathbf{l}}_l M^{\mathbf{itfl}}_{ijkl}}_{\text{input nonlinearity model}} \underbrace{(c_2 + \mathsf{Context}(i-j+1,k))}_{\text{contextual modulation}} \tag{16}$$

$$= c + \sum_{jkl} w^{\mathbf{t}}_j w^{\mathbf{f}}_k w^{\mathbf{l}}_l M^{\mathbf{itfl}}_{ijkl} \left(c_2 + \sum_{mnp} w^{\boldsymbol{\tau}}_m w^{\boldsymbol{\phi}}_n w^{\boldsymbol{\lambda}}_p [M^{\boldsymbol{\tau\phi\lambda}}(i-j+1,k)]_{mnp} \right) \tag{17}$$

To render this in the standard multilinear form we define a 7-dimensional array $\mathbf{Q}^{\mathbf{itfl}\boldsymbol{\tau\phi\lambda}}$:

$$Q^{\mathbf{itfl}\boldsymbol{\tau\phi\lambda}}_{ijklmnp} = \begin{cases} M^{\mathbf{itfl}}_{ijkl}[M^{\boldsymbol{\tau\phi\lambda}}(i-j+1,k)]_{mnp} & (j,k,l,m,n,p) \le (J,K,L,M,N,P) \\ 1 & (j,k,l,m,n,p) = (J+1,K+1,L+1,M+1,N+1,P+1) \\ 1 & (j,k,l) \le (J,K,L); \quad (m,n,p) = (M+2,N+2,P+2) \\ 0 & \text{otherwise} \end{cases}$$

Then, with $\mathbf{w}^{\mathbf{t}} \ldots \mathbf{w}^{\boldsymbol{\lambda}}$ correspondingly augmented with 1 or 2 additional dimensions each, we obtain the expression

$$\widehat{\mathbf{r}} = (\mathbf{w}^{\mathbf{t}} \otimes \mathbf{w}^{\mathbf{f}} \otimes \mathbf{w}^{\mathbf{l}} \otimes \mathbf{w}^{\boldsymbol{\tau}} \otimes \mathbf{w}^{\boldsymbol{\phi}} \otimes \mathbf{w}^{\boldsymbol{\lambda}}) \bullet \mathbf{Q}^{\mathbf{itfl}\boldsymbol{\tau\phi\lambda}}$$

with the constants in (17) recovered as

$$c = w^{\mathbf{t}}_{J+1} w^{\mathbf{f}}_{K+1} w^{\mathbf{l}}_{L+1} w^{\boldsymbol{\tau}}_{M+1} w^{\boldsymbol{\phi}}_{N+1} w^{\boldsymbol{\lambda}}_{P+1}$$
$$c_2 = w^{\boldsymbol{\tau}}_{M+2} w^{\boldsymbol{\phi}}_{N+2} w^{\boldsymbol{\lambda}}_{P+2}$$

Note that the parameter vectors $\mathbf{w}^{\boldsymbol{\tau}}, \mathbf{w}^{\boldsymbol{\phi}}, \mathbf{w}^{\boldsymbol{\lambda}}$ associated with the contextual subunit are independent of the time-lag j and the frequency k. As such, the action of the contextual subunit may be interpreted as an adjustment of the stimulus that happens prior to the action of an input nonlinearity model, and independent of the frequency. However, the parameters determining the contextual effects are estimated simultaneously with the others; algorithmically, there is no hierarchy.

We may re-group some of the parameter vectors as before – for example, assume all components are separable, or group time and sound level together so that the parameters are $\mathbf{w}^{\mathbf{tl}} \otimes \mathbf{w}^{\mathbf{f}} \otimes \mathbf{w}^{\boldsymbol{\tau}} \otimes \mathbf{w}^{\boldsymbol{\phi}} \otimes \mathbf{w}^{\boldsymbol{\lambda}}$.

Since the different components of the models interact multiplicatively, there is an ambiguity of scale in the parameters. For example, if $\mathbf{w^l}$ were multiplied by a constant factor and $\mathbf{w^t}$ divided by that same factor, an identical model would result. We resolve this ambiguity in the input nonlinearity part of the model by rescaling all parameter vectors but the first so that the maximal absolute-value of the vector components is 1. The first parameter vector then has an unconstrained, but well-defined, y-axis, with units of spikes/second. The context part of the model can be separately rescaled relative to c_2; we resolve this by dividing by this constant so that $c_2 = 1$, and then scaling $\mathbf{w^\phi}$ and $\mathbf{w^\lambda}$ to have maximal absolute-values of 1, thereby assigning the scale of the context terms to $\mathbf{w^\tau}$.

The dimensionalities used in this paper are as follows:

- $\mathbf{w^t}$: t ranged from 0 to 10 (in units of 20 ms time bins), i.e. $J = 11$.
- $\mathbf{w^f}$: there were either 24 or 48 frequency bins in the stimuli, so that $K = 24$ or 48.
- $\mathbf{w^l}$: the stimuli contained 10 sound levels, so $L = 10$.
- $\mathbf{w^\tau}$: the contextual subunit was defined to stretch from 0 to 10 time bins into the past, so that $M = 11$.
- $\mathbf{w^\phi}$: the contextual subunit was defined to stretch 5/12 octaves to either side of a target tone (or less, if the target tone appeared near to the spectral boundary of the stimulus). Since the frequency bands in the stimulus were 1/12 octaves apart, $N = 11$.
- $\mathbf{w^\lambda}$: the stimuli contained 10 sound levels, so $P = 10$.

Results

Neural recordings

The neuronal population consisted of 147 units, 97 from mice and 50 from rats. Of these, 39 recordings (21 in rats and 18 in mice) were made under stimulation by higher frequency sounds (25-100 kHz), and 108 (76 in mice and 32 in rats) were made under stimulation by lower frequency sounds (2-32 kHz). In mice, about half of the recordings came from A1, and the other half came from AAF; all the recordings in rats were from A1. In all the recordings used for analysis, an estimate of the repeatable stimulus-locked *signal power* (Sahani and Linden, 2003b) was at least one standard deviation greater than zero. Results based on subsets of these data have been reported previously (Linden et al., 2003; Sahani and Linden, 2003a).

Performance of the models

The predictive performance of the STRF and multilinear models fit to each recording in the population was evaluated by cross-validation. The 60 s stimulus presentation time was first divided into ten segments of equal size. The models were then trained using data from nine of these segments and tested on the remaining one; this procedure was repeated ten times, once for each of the ten possible test segments. The predictive power of a model was defined to be the average predictive power on the ten cross-validation datasets (Sahani and Linden, 2003b). Predictive power is a measure of performance that takes into account trial-to-trial variability of the neuronal response; its expected value for the "perfect" model is 1, and for a model predicting only the mean firing rate it is 0. Fig. 3 shows these predictive powers for the fully separated input nonlinearity and context models, compared to a well-fit STRF model (estimated by automatic smoothness determination; Sahani and Linden, 2003a).

The predictive power obtained by cross-validation is a *lower bound* on the true predictive power achievable by the model. As finite data are used for training, the estimated parameters are inevitably overfit to the training

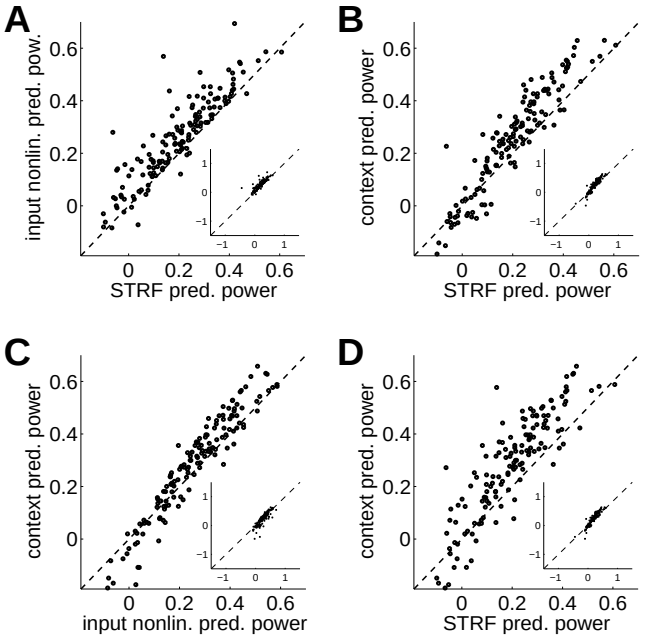

Figure 3: Lower bounds on predictive power for neurons in rodent auditory cortex. ***A***: predictive power of the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ input nonlinearity model versus that of a well-fit linear (STRF) model. ***B***: predictive power of the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}} \otimes \mathbf{w}^{\tau} \otimes \mathbf{w}^{\phi} \otimes \mathbf{w}^{\lambda}$ context model versus that of the linear model. ***C***: predictive power of the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}} \otimes \mathbf{w}^{\tau} \otimes \mathbf{w}^{\phi} \otimes \mathbf{w}^{\lambda}$ context model versus that of the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ input nonlinearity model. Negative predictive power means that the prediction is worse than predicting a constant mean firing rate. The input nonlinearity model performs better on most neurons than the STRF model, and the context model performs better than both on the majority of neurons, despite having the largest number of parameters. ***D***: predictive power of the $\mathbf{w^{t}} \otimes \mathbf{w^{f}} \otimes \mathbf{w^{l}} \otimes \mathbf{w}^{\tau} \otimes \mathbf{w}^{\phi} \otimes \mathbf{w}^{\lambda}$ context model versus that of the STRF model. The cross-validation performance of the fully separated context model is higher than that of all the other models.

data, which leads to suboptimal predictive performance on the test segment. (Negative predictive powers on test data are also possible: this means that the model prediction is less accurate than a constant prediction of the mean firing rate would be.) A complementary *upper bound* on the predictive power can be obtained by fitting the *unregularized* model to the entire dataset and computing the resultant predictive power directly on the training data. This overestimates the true value, because the modelling of noise in the training data cannot be distinguished from accurate prediction. The "true" predictive power of the model, i.e., the predictive power it would have when trained on an infinitely large dataset, lies between these upper and lower bounds (Sahani and Linden, 2003b). These bounds become tighter as the trial-by-trial variability of the response, the *noise power*, decreases. Thus, following Sahani and Linden (2003b), we consider the population as a whole, and extrapolate both the upper and lower bounds to the point of zero noise using polynomial fits, whose order is chosen by leave-one-out cross-validation. This yields relatively tight bounds on the performance expected of the model if it were applied to a hypothetical neuron drawn from a similar population but with a completely stimulus-determined response. The extrapolation is shown in Fig. 4.

Using this analysis, we find that the predictive power of the input nonlinearity model $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ (0.27 – 0.40, expressed as a fraction of the predictable signal power) is modestly larger than that of the STRF model (0.23 – 0.37). The predictive power of the context model is substantially higher (0.32 – 0.52), with the midpoint between the bounds being about 1.4 times that of the STRF model. This finding demonstrates that taking into account local acoustic context (based on a spectrographic window stretching over recent history and nearby

frequencies) through the structure of the context model allows the multilinear description to capture more of the dynamic behaviour of auditory cortical neurons than the linear STRF-based descriptions.

Fig. 5 shows four examples of how the predictions for the STRF and the context models differ. The context model generally predicts the PSTH more accurately, especially at stimulus-evoked peaks in the firing rate.

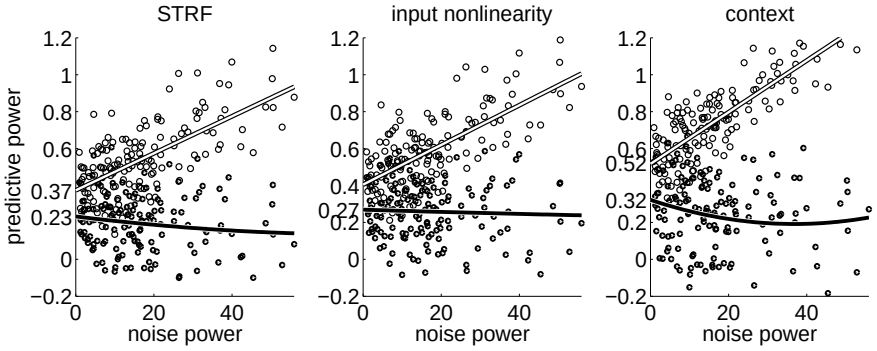

Figure 4: Interpolation of the predictive power to the zero noise limit for the pooled population. *Open circles and white line*: upper bound predictive powers. *Closed circles and black line*: lower bound predictive powers. The input nonlinearity model is the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ model and the context model is the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}} \otimes \mathbf{w^{\tau}} \otimes \mathbf{w^{\phi}} \otimes \mathbf{w^{\lambda}}$ model (see Methods). Note that both bounds for the multilinear models might be underestimated because their objective function may have multiple local minima. Nevertheless the input nonlinearity model performs better than the STRF model and the context model performs better than both. All three models sometimes predict negative firing rates; when those negative rates are set to zero, the predictive power bounds of the models increase by between 1 and 2 percent.

Input nonlinearity model

Sensitivity to sound level. We used the input nonlinearity model $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ to study the sensitivity of neuronal responses to sound level within the DRC spectrogram. Fig. 6A shows the model parameters and predictive performance for one such model. The inferred level-sensitivity parameter, $\mathbf{w^{l}}$, exhibits a mild threshold and remains fairly linear (in logarithmic terms) above about 45 dB SPL. Note that this inferred level-dependence may not be the same as the usual rate-level function, for two reasons. First, the rate-level function is conventionally measured using isolated tones presented in silence. Level sensitivity in the context of a complex sound may be quite different Dean et al. (2005). Second, the estimate of $\mathbf{w^{l}}$ is made in the context of subsequent integration through the time-frequency weights $\mathbf{w^{tf}}$. Estimates of the two parameter vectors are interdependent. Similarly, the weights $\mathbf{w^{tf}}$ learnt through this model are sometimes different from those of the STRF (Fig. 6E).

Fig. 6B shows some more examples of level-sensitivity profiles $\mathbf{w^{l}}$ learnt using different neuronal responses. These are illustrative of the variety of shapes seen in the population. To summarise the observed distribution, 5 prototypical profiles were designed by hand, and each learnt profile classified according to the prototype with which its dot-product was greatest. The results are shown in Fig. 6C. Note that the profiles were chosen to illustrate the range of shapes observed; we did not see evidence of clustering in the data.

Inseparabilities. As described in Methods, the parameters of the input nonlinearity model may be grouped in several different ways. These groupings can be used to assess several forms of inseparability that may be

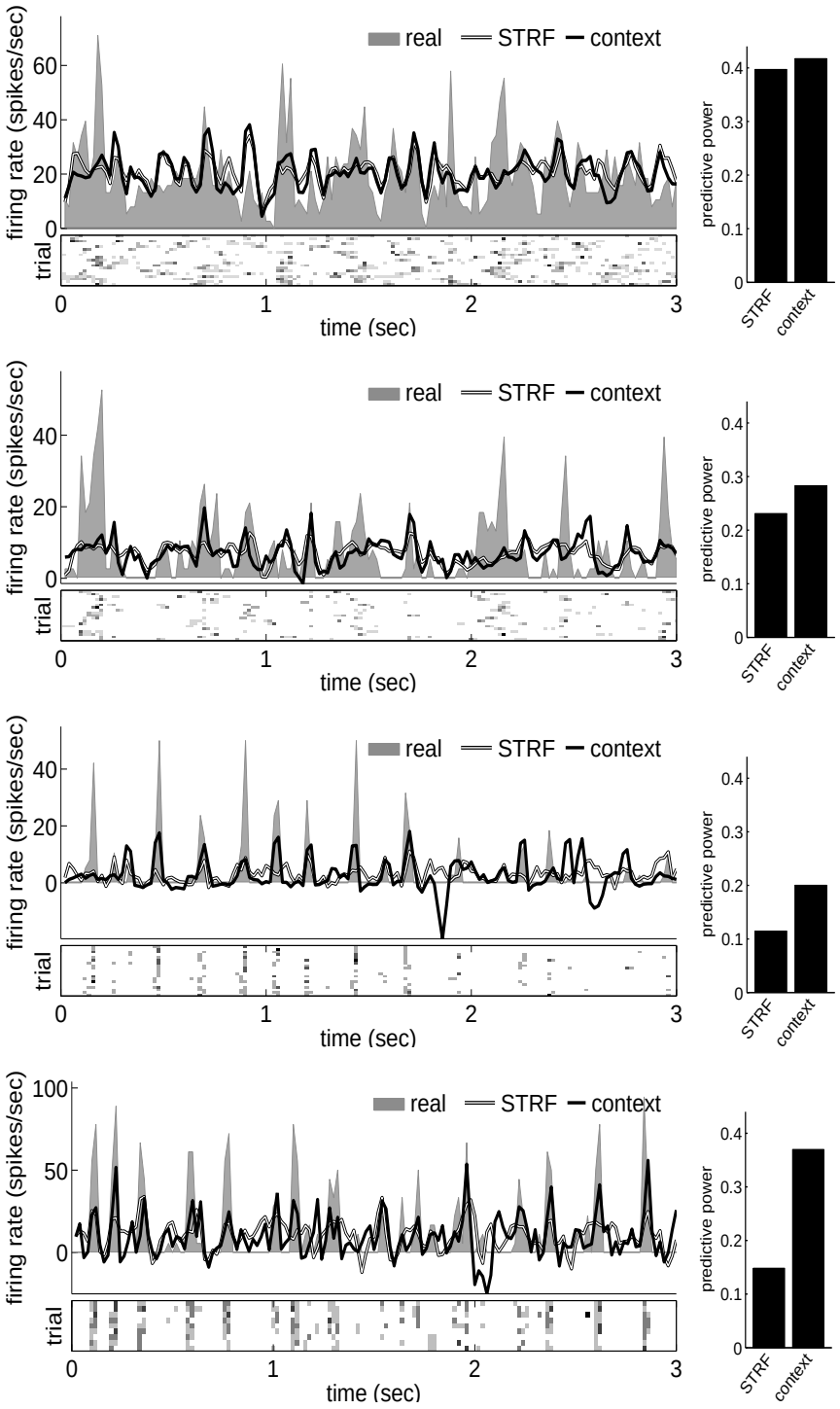

Figure 5: PSTH and cross-validation predictions for four example cells which illustrate a range of predictive power increments of the context model over the STRF model. Within each panel, *top:* real PSTH (*gray*), STRF and context model predictions (*white and black*); *bottom:* spike rate during each trial. Grayscale indicates number of spikes in each 20 ms bin. On the *right* are shown the cross-validation predictive powers. The context model generally predicts the PSTH more accurately than the STRF model, especially at the peaks of the PSTH (e.g. *bottom panel*). The *top panel* shows a case in which the context model's prediction quality is roughly the same as that of the STRF model.

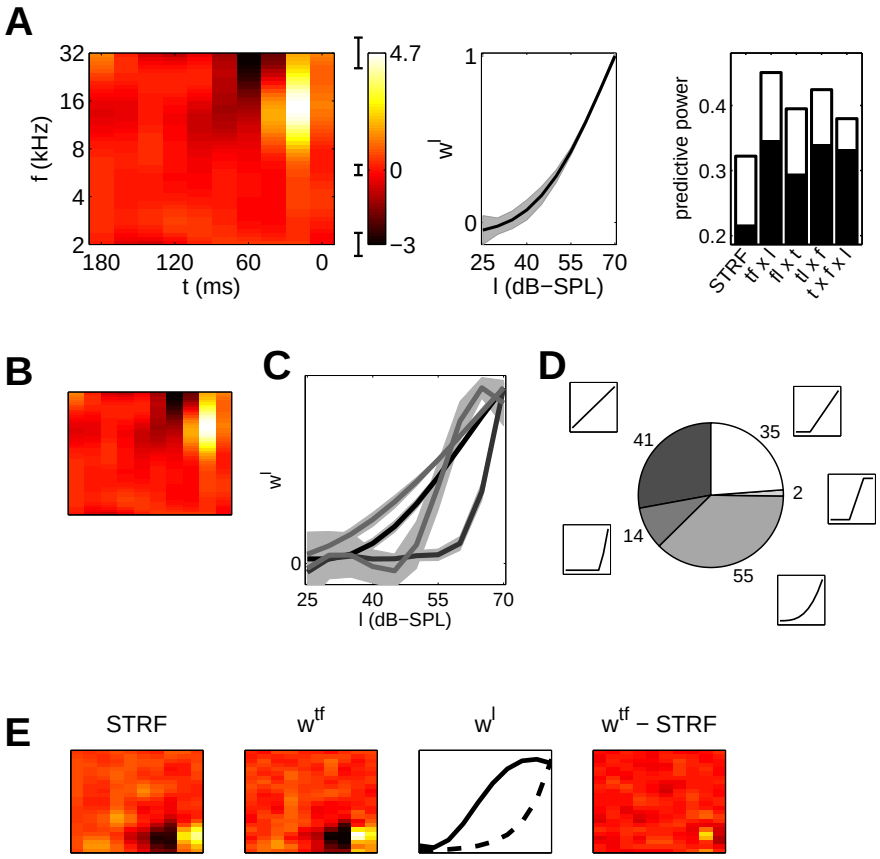

Figure 6: ***A:*** Input nonlinearity model $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ for a neuron in mouse auditory cortex. *Left:* the time-frequency component $\mathbf{w^{tf}}$ of the model. The average standard error of the weights of $\mathbf{w^{tf}}$ across time-frequency bins is shown by the error bar at the 0 position of the color bar. The standard error of the peak weight is shown at the top of the bar; the error at the minimum is shown at the bottom of the bar. Units are spikes/sec. *Center:* the sound level component of the model. The one standard deviation error bars are shown as a grey area. All error bars were calculated by bootstrap methods. This graph has been normalized and thus has no units. *Right:* predictive power of the STRF model and all the possible configurations of the input nonlinearity model. The white bars show the predictive power of the regularized models on the training data, and the black bars show the predictive power on cross-validation data. For this neuron, the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ model has the highest predictive power, and indeed, the $\mathbf{w^{tf}}$ receptive field can be seen to be inseparable, because the peak excitation (at a 20 ms lag time) occurs at a lower frequency than the peak suppression (at a 60 ms lag time). ***B:*** For comparison, the STRF estimated on the same data, which in this case differs little from $\mathbf{w^{tf}}$. ***C:*** Various $\mathbf{w^{l}}$ from the population. Grey areas indicate one standard error. ***D:*** Number of cells whose $\mathbf{w^{l}}$ fits are best approximated by each of five idealized shapes. The shapes were chosen by hand to reflect the variety of shapes of $\mathbf{w^{l}}$ derived from the data. Similarity between the actual $\mathbf{w^{l}}$ and the idealized shape is defined to be the normalized dot product between the graphs. No non-monotonic shapes were observed, but this may be a result of the sound level range used in the experiments, or of the inequivalence between $\mathbf{w^{l}}$ and the rate-level function. The smoothed ramp shape was the most common. ***E:*** Example of a cell for which $\mathbf{w^{tf}}$ of the input nonlinearity model is different from the STRF. The difference is shown in the fourth panel; the color code is the same in all panels. The third panel shows $\mathbf{w^{l}}$ for the input nonlinearity model (*solid*), and the fixed input scaling function that is assumed for all STRF models (*dashed*). (This function, which is linear in sound pressure, was found to produce better STRF results across the population of cells than a linear function in dB.) The $\mathbf{w^{tf}}$ receptive field shows a peak response at a larger lag time than suggested by the STRF. The predictive power of the input nonlinearity model was 1.6 times that of the STRF model, on both training and cross-validation data.

present in the stimulus-response functions. Time-frequency inseparabilities in the STRF have been studied extensively (e.g. Kowalski et al., 1996a,b; Depireux et al., 2001; Linden et al., 2003) and related properties (such as temporal symmetry; Simon et al., 2007) have been documented. The opportunity exists to extend such studies to new forms of interaction.

The optimal form of inseparability for each neuron was assessed by comparing the cross-validation predictive powers of the variously grouped models. Thus, a cell would be classified as having a frequency-level inseparable receptive field if the $\mathbf{w^{fl}} \otimes \mathbf{w^{t}}$ model performed better on cross-validation data than the $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$, $\mathbf{w^{tl}} \otimes \mathbf{w^{f}}$, $\mathbf{w^{t}} \otimes \mathbf{w^{f}} \otimes \mathbf{w^{l}}$ and the STRF models. There are two caveats to this classification. First, more than one inseparable model could perform better than the fully separated one, suggesting that all three stimulus dimensions might interact in that neuron's response function. As the full model (with no separated parameters) could not be fit reliably with the data available we did not explore this three-way interaction further. The second caveat is that the different models have different numbers of parameters, and are thus differently prone to overfitting. Thus, while the fully separated $\mathbf{w^{t}} \otimes \mathbf{w^{f}} \otimes \mathbf{w^{l}}$ model performs best on cross-validation data for many neurons (Fig. 7C) it is not possible to say whether this reflects genuine separability, or simply greater reliability in the estimates of a smaller number of parameters.

With these caveats, Fig. 7C shows the proportion of neurons best described by each of the configurations of the input nonlinearity model. These results suggest that frequency-level and time-level interactions, as well as the more widely studied time-frequency interactions, are present in the responses of many neurons to complex sounds.

The $\mathbf{w^{tf}} \otimes \mathbf{w^{l}}$ model. Separabilities in time and frequency have been extensively documented. The input nonlinearity model also recovers such inseparabilities in a number of neurons in the population (Fig. 7C). For these neurons, then, the form of frequency integration varies with lag time.

The $\mathbf{w^{fl}} \otimes \mathbf{w^{t}}$ model. We also observed several cells with frequency-level inseparabilities (Fig. 7C). Such inseparabilities are consistent with previous reports of frequency response areas derived using isolated tone pips, which can have quite complex and asymmetric shapes. A frequency-level inseparability indicates that sound level is processed differently for different frequencies; or alternatively, that frequency integration depends on sound level. Fig. 7A shows an example of such a cell.

The $\mathbf{w^{tl}} \otimes \mathbf{w^{f}}$ model. Finally, we found a substantial number of cells whose optimal inseparability was in time and sound level (Fig. 7C). This form of inseparability indicates that during the processing of a complex stimulus, sound level integration depends on lag time. Fig. 7B shows an example of such a neuron. The time-level receptive field is clearly inseparable, showing a fast response to loud sounds, without any suppression. Prolonged suppression occurs only for sounds of *intermediate* loudness. In other words, the response to sound is monotonic at short lag times, but non-monotonic at larger lag times. Thus, for this cell, the temporal processing of sounds is critically dependent on the sound level.

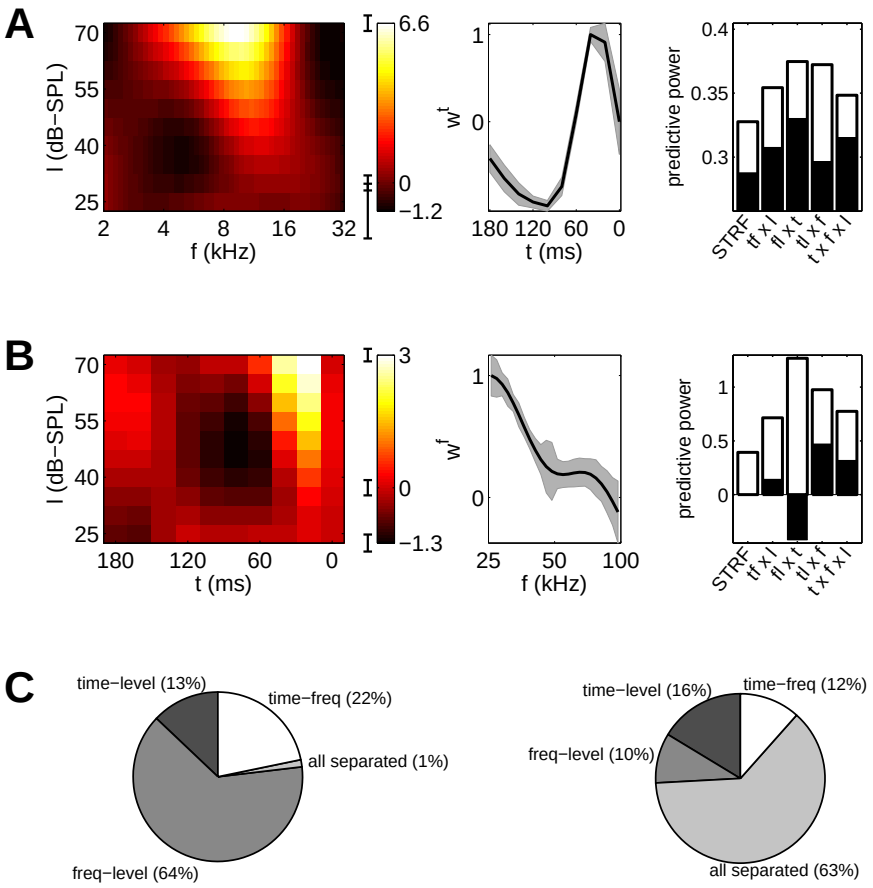

Figure 7: ***A:*** Input nonlinearity model $\mathbf{w^{fl}} \otimes \mathbf{w^{t}}$ for a neuron in mouse auditory cortex. Conventions as in Fig. 6. This unit was classified as having a frequency-level inseparable receptive field. ***B:*** Input nonlinearity model $\mathbf{w^{tl}} \otimes \mathbf{w^{f}}$ for another auditory cortical neuron in the mouse. Note that the cross-validation predictive power (e.g. second black bar) can be negative, meaning that the model does worse at predicting the cell's response than a constant average firing rate prediction would. This unit was classified as time-level inseparable on cross-validation data. ***C:*** Proportions of the population of rodent auditory cortical neurons best described by each of the possible configurations of the input nonlinearity model. *Left:* Proportions for training data. *Right:* Proportions for testing data. The discrepancy between the two pie charts illustrates one of the difficulties in classifying the neurons into groups of optimal inseparability, which is that models with more parameters are more likely to overfit to noise in training data, and thereby underfit testing data. Importantly, however, all of time-frequency, time-level and frequency-level groupings occur in substantial proportions in both pie charts, showing that all three forms of inseparability exist in auditory cortex.

The context model

The context model (see Methods and Fig. 2) can be viewed as combining two input nonlinearity models. One (A in Fig. 2) determines the modulation of the effective sound level of a time-frequency element of the stimulus by its local context; the other (B in Fig. 2) determines the predicted firing rate using these modulated values. The algorithm used to estimate the parameters fits both models simultaneously, allowing the parameters (A) to affect the estimate of the parameters (B) and *vice versa*.

More specifically, the context model extends the input nonlinearity model through the term $\mathbf{w^{\tau}} \otimes \mathbf{w^{\phi}} \otimes \mathbf{w^{\lambda}}$.

Thus, the effective intensity of a time-frequency element (i.e., the stimulus power $s(i,k)$ in some time bin i and frequency bin k) is modulated by each time-frequency element that precedes it (within a window of times and frequencies) in a way that depends on their separation in time through $\mathbf{w}^{\boldsymbol{\tau}}$, their separation in frequency through $\mathbf{w}^{\boldsymbol{\phi}}$ and on the level of the modulating time-frequency element through $\mathbf{w}^{\boldsymbol{\lambda}}$.

The context effect. Since the emphasis of this section is on the form of the context terms, only fully separated context models are shown in Fig. 8. As illustrated in that figure, $\mathbf{w}^{\boldsymbol{\phi}}$ is negative across most frequency differences, while $\mathbf{w}^{\boldsymbol{\tau}}$ and $\mathbf{w}^{\boldsymbol{\lambda}}$ are generally positive. Thus the context term $\mathbf{w}^{\boldsymbol{\tau}} \otimes \mathbf{w}^{\boldsymbol{\phi}} \otimes \mathbf{w}^{\boldsymbol{\lambda}}$ is generally suppressive over most time-frequency differences.

The $\mathbf{w}^{\boldsymbol{\tau}}$ parameters illustrated show their greatest suppression for lags of 20-100 ms, depending on the neuron. The $\mathbf{w}^{\boldsymbol{\phi}}$ values show greatest suppression when the modulating tone is at the same frequency as its target. This suppression falls off with greater frequency separation, vanishing at a separation of about half an octave (or more in some cells; data not shown). The sound level representation of the context $\mathbf{w}^{\boldsymbol{\lambda}}$ is generally monotonic, and sometimes different from $\mathbf{w}^{\mathbf{l}}$.

Fig. 8C demonstrates the context model for one neuron, for which context tones just before the target tone can have an excitatory effect ($w^{\boldsymbol{\tau}}$ is negative at $\tau = 20$ms; with $\mathbf{w}^{\boldsymbol{\phi}}$ generally negative this together implies excitation). Fig. 8A, which displays the context model parameters for a different neuron, shows two slightly excitatory humps at the edges of $w^{\boldsymbol{\phi}}$. This structure of $\mathbf{w}^{\boldsymbol{\phi}}$ was also observed for other neurons, especially in a few cases when the range of ϕ was extended (data not shown).

The context model and STRF models. Suppressive stimulus effects can be captured by linear models though negative regions in their temporal or spectrotemporal receptive fields. The context terms of the context model provide an additional way to create suppressive effects. We wondered if the suppressive regions seen in STRFs might be better accounted for by the contextual component of the context model than by negative weights in $\mathbf{w}^{\mathbf{tf}}$. To measure this, we fit multilinear models with context terms (that is, the fully separated context model) and without context terms (the fully separated input nonlinearity model) to the data, and examined the relative amount of suppression as reflected by $\mathbf{w}^{\mathbf{t}}$, for each neuron in the population. Fig. 9 shows the relative suppression $\min(\mathbf{w}^{\mathbf{t}})/(\max(\mathbf{w}^{\mathbf{t}}) - \min(\mathbf{w}^{\mathbf{t}}))$ for both models. There is a marked decrease in the relative amount of observed suppression when the contextual terms are present, and sometimes the suppression disappears. This suggests that the suppressive regions seen in the STRFs are, at least in part, due to contextual effects, rather than to additive suppression.

Fig. 10 illustrates a mechanism that may be responsible for this effect. The PSTH of a hypothetical neuron, modelled by a context model (panel A) responding to a stimulus similar to the DRC used in the experiments, was fit with both an STRF model (panel B, top) and a context model (dashed line, panel A; and panel B, bottom). The original context model contained no additive suppression, as there is no trough in $\mathbf{w}^{\mathbf{t}}$, but it does contain multiplicative stimulus interactions (through $\mathbf{w}^{\boldsymbol{\tau}}$ and $\mathbf{w}^{\boldsymbol{\phi}}$). The fitted STRF model, on the other hand, contains a clearly visible suppressive region following the excitatory peak. This again shows that the suppressive regions in STRFs fit to experimental data may reflect multiplicative stimulus interactions (forward suppression, two-tone interactions) rather than genuine additive suppression.

Simulation of STRF features changing with spectral density. STRFs fit to data collected using different stimulus classes are generally not the same (e.g. Theunissen et al., 2000; Valentine and Eggermont, 2004; see also Rotman et al., 2001; Bar-Yosef et al., 2002; Woolley et al., 2006). We used the context model to help understand certain systematic stimulus-dependent changes in STRFs related to the spectral density of a

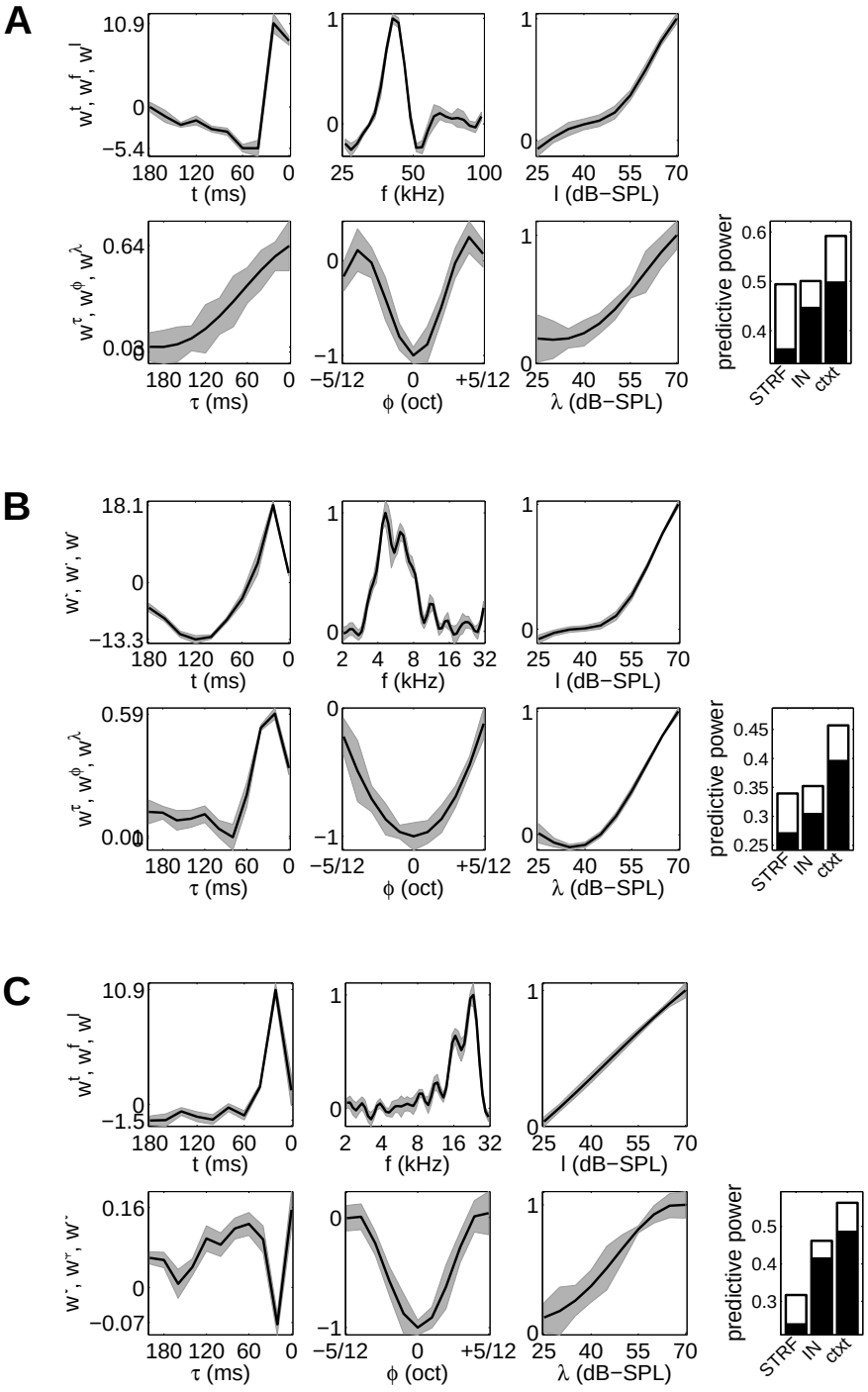

Figure 8: ***A:*** A fully separated context model for a rat auditory cortical neuron. Shaded regions indicate one standard error. The x-axes of the top row are as in the input nonlinearity model: lag time t, frequency f, and sound level l. Units of $\mathbf{w^t}$ are spikes/sec; $\mathbf{w^f}$ and $\mathbf{w^l}$ have no units as they have been normalized. The bottom row shows the context components, with the following axes: time difference between context and target tones τ, frequency difference between the tones ϕ, and sound level of the context tone λ. These components have no units but their scale relative to $\mathbf{w^t}$ is contained in the values of $\mathbf{w^\tau}$, as $\mathbf{w^\phi}$ and $\mathbf{w^\lambda}$ have been normalized. The bottom right panel displays the predictive powers of the STRF model, the fully separated input nonlinearity model, and the context model shown. The white and black bars show the predictive power on training data and cross-validation data, respectively. ***B:*** A context model for a mouse auditory cortical neuron. The context effect follows a faster time course than the example in ***A***. ***C:*** A context model for a rat auditory cortical neuron. In this case, the context could have a facilitatory effect on the target time-frequency element. Such behaviour was relatively rare in our population.

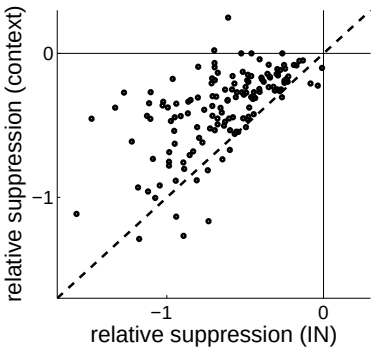

Figure 9: Depth of trough of $\mathbf{w^t}$ of the fully separated input nonlinearity and context models. Values on the axes are $\min(\mathbf{w^t})/(\max(\mathbf{w^t}) - \min(\mathbf{w^t}))$, for the input nonlinearity model and for the context model.

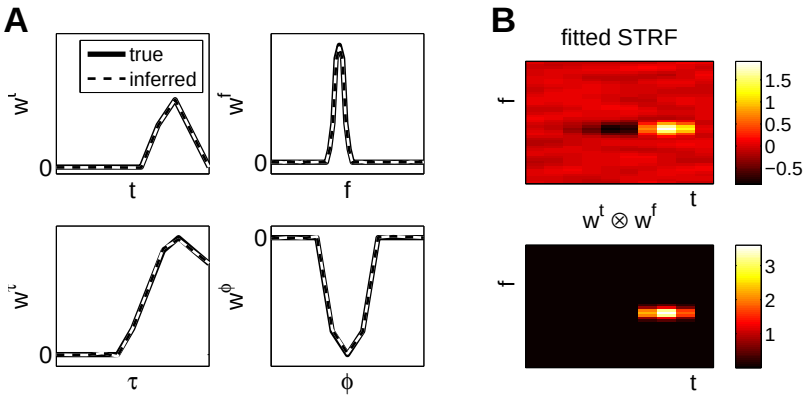

Figure 10: Fitting a linear STRF model to a hypothetical nonlinear neuron described by a context model. ***A:*** the original context model (*black*) and the context model fit to the simulated PSTH (*dashed*). The fitted model matches the original very well. The fitting was done without regularization; the simulated PSTH was taken to be the noiseless output of the original context model, thresholded to avoid negative firing rates. The parameters $\mathbf{w^l}$ and $\mathbf{w^\lambda}$ were set to be linear and were not fit for brevity. ***B, top:*** an STRF model fit to the simulated PSTH, ***B, bottom:*** the $\mathbf{w^{tf}}$ receptive field of the hypothetical neuron given by $\mathbf{w^t} \otimes \mathbf{w^f}$. The additive part $\mathbf{w^t} \otimes \mathbf{w^f}$ of the original context model does not show additive suppression, whereas the STRF does show a suppressive region. This suppressive region in the STRF is therefore purely due to the contextual effects of $\mathbf{w^\tau} \otimes \mathbf{w^\phi}$.

sound. The dependence of the STRF on the spectral density of a DRC can be summarized as follows (Blake and Merzenich, 2002): a higher spectral density leads, on average, to a lower excitatory peak in the STRF, and in some cells, the suppressive region only appears at higher spectral densities. Could these experimental results be explained by the context model? We used the context model of Fig. 8B (fit on real data at a single spectral density) with five stimuli of varying spectral density, to generate five *simulated* PSTHs. Then we fit STRF models for these five stimuli and simulated PSTHs. This process is summarized in Fig. 11A.

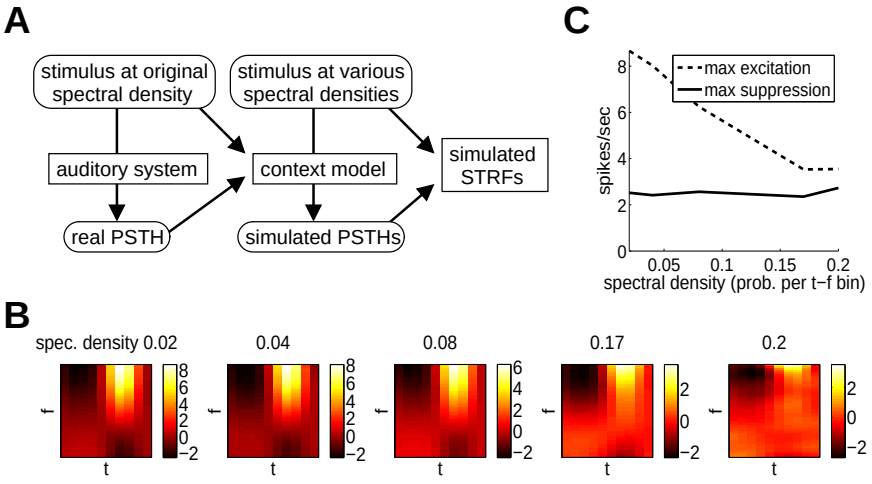

Figure 11: ***A:*** Summary of how STRFs were fit to the outputs of a context model. The context model was trained on experimental data that was gathered using a stimulus at a single spectral density, then run on stimuli of various spectral densities to generate simulated PSTHs. ***B:*** STRF fits to simulated PSTHs from context models that were run on stimuli of increasing spectral density. The peak excitation decreases monotonically, and so does bandwidth. Both of these observations were also made experimentally. ***C:*** The peak excitation and the magnitude of the maximum suppression of the STRFs shown in ***B***. The peak excitation decreases monotonically whereas the suppression stays roughly constant, in accordance with experimental observations (Blake and Merzenich, 2002).

The five STRFs show phenomena similar to those observed experimentally, as shown in Fig. 11B. First, there is spectral narrowing at higher spectral densities. Second, the excitatory peak becomes lower at higher spectral densities, whereas the suppressive depth stays roughly constant, as summarized in Fig. 11C. This graph is qualitatively the same (a monotonic decrease in peak value, and a roughly constant suppression) as the experimental observations of Blake and Merzenich (2002). Thus, the context model is a candidate for explaining the changes in STRFs at various stimulus densities.

Besides the behaviour shown in Fig. 11, there are context models (both fit to real data and constructed by hand) that can be used to generate data for which STRFs would have other forms of dependence on the spectral density of the stimulus. For example, the STRFs may show no suppressive region at low spectral density, and a deepening suppressive region at higher spectral density, which is similar to the experimentally observed appearance of suppressive regions at higher spectral density. There are also context models for which STRF fits would change very little with spectral density. Cells with similar response properties have also been reported experimentally.

Discussion

We have constructed a class of models that describe the transformation between a sound and the firing rate of a neuron in the auditory cortex. These models were used here to analyze responses to DRC stimuli, but are applicable to the analysis of responses to arbitrary sounds. The models make possible novel analyses of auditory cortical responses to complex sounds, and also unify measurement of many different response properties that were previously analyzed separately, such as the rate-level function, the STRF, and contextual effects.

What is gained by this unification? Why should we not, for instance, measure the rate-level function and the STRF separately instead of estimating the input nonlinearity model? First, the stimulus conditions under which rate-level functions are typically measured (tones in silence) are different from those used for STRF estimation (continuous complex sounds); the state of the auditory system is thus likely to be different in the two cases. The present approach allows us to estimate sensitivities to different stimulus dimensions simultaneously, from a single (complex) stimulus, thus obtaining a more coherent picture of the function of the auditory system in a single context. Second, the terms in the models depend on one another. For example, a complex sound contains energy at different sound levels. An STRF fit to such data carries with it an implicit model for sound level sensitivity; this model is typically the identity function, and so entries of the STRF are weighted proportional to sound level (in, say, dB SPL). To the extent that this assumption is not the most appropriate, the STRF estimate will be affected. Estimating neuronal sound level sensitivity directly, using the input nonlinearity model, produces models that describe the system more accurately, as demonstrated by the improvement in model predictive power, and the weights $\mathbf{w}^{\mathbf{tf}}$ in the input nonlinearity model are sometimes not equal to those in the STRF. Similarly, for time-level inseparable cells, fitting a time-level receptive field would be impossible without a model for frequency integration; poor with a suboptimal model of it; and not general if one made frequency integration irrelevant by using a fixed spectral content in the experiment. A third and final example of the benefits of a unified model, as demonstrated in Results, is that the context terms of the context model have important effects on the inhibitory region of the spectrotemporal receptive field — something that would not be evident if an STRF and a two-tone receptive field were measured separately.

The input nonlinearity. Finding an appropriate single input nonlinearity for all neurons can raise the average predictive power of an STRF model (Gill et al., 2006). However, our data revealed various shapes of $\mathbf{w}^{\mathbf{l}}$. Thus estimating the input nonlinearity adds neuron-specific detail and provides a better understanding of how populations of neurons respond to level in complex sounds.

Separability of receptive fields. Using the input nonlinearity framework, we found frequency-level and time-level inseparabilities during processing of complex sounds in significant proportions of our neuronal population. Attention has previously focused on time-frequency inseparabilities (Kowalski et al., 1996a; Depireux et al., 2001; Miller et al., 2002; Linden et al., 2003; Simon et al., 2007). The present models open up the possibility of exploring the properties of the other two types of inseparable receptive field.

Time-frequency inseparable receptive fields may underlie frequency sweep direction selectivity (deCharms et al., 1998). The inseparabilities in the other pairs of variables may reflect other functions. For example, cells with a time-sound level inseparable receptive field may respond to sounds with increasing or decreasing levels, which feature, for instance, in vocalisations (particularly speech). Without multilinear models, such effects are very hard to study in the context of complex sounds.

Stimulus context. Including local stimulus context allows us to predict auditory cortical responses more accurately than before. The context model should also be applicable to pre-cortical auditory structures, where contextual influences of the sort described first arise.

The context terms reveal forward masking suppression in auditory cortical responses to complex sounds, lasting about 200 ms, with a frequency width of about half an octave in most cells. Some neurons also show forward excitation through positive humps at the edges of the frequency-difference term of the context model. These results are consistent with those of previous physiological studies using two-tone paradigms in A1 (Brosch and Schreiner, 1997; Calford and Semple, 1995), and with previous psychophysical studies (Jesteadt et al., 1982; Moore, 1980, 1997). The dominance of forward suppression over forward excitation is consistent with recent subthreshold studies (Wehr and Zador, 2005). Again, the current models show the impact of such effects in the responses to complex sounds. Longer time scales of adaptation also exist in A1 (McKenna et al., 1989; Ulanovsky et al., 2003; Bartlett and Wang, 2005) but were not studied here.

We also find that the context model has implications for the interpretation of results of previous STRF studies. Most notably, (1) the suppressive region of $\mathbf{w^t}$ is reduced when contextual effects are modeled explicitly (Fig. 9), and (2) the context model can account qualitatively for previous observations of stimulus-density-dependent STRFs, such as spectral narrowing and the changing ratio between the maximum and minimum values of the STRF (Fig. 11). These two observations are linked, as the average stimulus power in the context of a time-frequency element increases with spectral density. Thus, with increasing spectral density of the stimulus, the impact of the contextual subunit on the PSTH increases, and therefore the difference between a fitted STRF and the $\mathbf{w^t} \otimes \mathbf{w^f}$ component of a context model will also increase. One way in which the STRF may depart from $\mathbf{w^t} \otimes \mathbf{w^f}$ is through the appearance of the deepening suppressive region (see Fig. 10), in line with observation (1) above; another way may be through an idealized version of the systematic changes found by Blake and Merzenich (2002), such as spectral narrowing, in line with observation (2).

Interestingly, Blake and Merzenich found that STRF models with a more pronounced suppressive region were on average less predictive than STRF models where no suppressive region was observed. If, as our results here suggest, suppressive regions in the STRF are often signatures of significant contextual effects, then this observation may simply reflect the greater nonlinearity of cells that show them. An explicitly nonlinear approach (such as the context model) may well not show such an effect. (Alternatively, deeper suppressive regions may be associated with lower firing rates, leading to noisier STRFs and lower predictive power.)

Thus, we suggest that the previously reported changes in STRF structure with stimulus spectral density arise as a result of linear approximation of nonlinear cortical response functions (Christianson et al., in press), that are better described by the context model.

The models. The input nonlinearity model is a form of Hammerstein cascade (Narendra and Gallman, 1966; Hunter and Korenberg, 1986); however its development in the multilinear setting, and related advances in estimation, are more recent (Ahrens et al., in press). We are unaware of previous discussions of alternative grouping of stimulus dimensions facilitated by the multilinear view. The higher-order nonlinear models (e.g. the context model) are novel. Multilinear regression models have not received extensive attention in the statistics literature (but see e.g. Paatero, 1999). The fully inseparable model $\widehat{\mathbf{r}} = \mathbf{w^{tfl}} \bullet \mathbf{M}$ is a Generalized Additive Model (Hastie and Tibshirani, 1999; Breiman and Friedman, 1985), but its parameters are too numerous to make it useable (Aertsen and Johannesma, 1980). Other multilinear methods have been useful in various settings in the statistics and machine learning setting (e.g. Vasilescu and Terzopoulos, 2005; Tenenbaum and Freeman, 2000; Harshman and Lundy, 1994).

Previously proposed nonlinear models for auditory processing (e.g. Fishbach et al., 2001) tend to be designed to mimic the known physiology of the auditory system and often do not allow for straightforward parameter estimation. The statistical ease of estimating multilinear models directly from neuronal data makes them more similar to models based on Volterra-Wiener expansions (Marmarelis and Marmarelis, 1978). However,

prohibitive amounts of data are typically needed to estimate the large number of parameters of the latter class of models, so that in general they can only be fit by restricting their parameter space or focussing on a single stimulus dimension (Young and Calhoun, 2005). The data requirements of other nonlinear approaches (Brenner et al., 2000) also make them difficult to apply to auditory data.

Other forms of the multilinear model, perhaps similar to the context model, should allow for the study of within-stimulus interactions and stimulus-specific adaptation in a variety of sensory systems (e.g. Ahrens et al., 2006). In the present case, multilinear methods provide a highly efficient description of auditory cortical response functions; note that the fully separated context model, though better able to predict neural responses, has fewer parameters than an inseparable linear STRF model. Multilinear models provide a firm statistical foundation for analysing nonlinear neuronal response properties; allow response parameters to be learnt directly from the data rather than being hand-tuned; and reveal a rich variety of input nonlinearities, inseparabilities, and context effects in auditory cortical responses to complex sounds, thereby introducing new possibilities for data analysis and extending existing understanding of auditory processing of complex sounds.

References

Aertsen AMHJ, Johannesma PIM (1980) Spectro-temporal receptive fields of auditory neurons in the grassfrog. I: Characterization of tonal and natural stimuli. Biol Cybern 38:223–234.

Aertsen AMHJ, Johannesma PIM, Hermes DJ (1980) Spectro-temporal receptive fields of auditory neurons in the grassfrog. II: Analysis of the stimulus-event relation for tonal stimuli. Biol Cybern 38:235–248.

Ahrens MB, Paninski L, Petersen RS, Sahani M (2006) Input nonlinearity models of barrel cortex responses. Abstract #212, 15th annual CNS meeting.

Ahrens MB, Paninski L, Sahani M (in press) Inferring input nonlinearities in neural encoding models. Network.

Bar-Yosef O, Rotman Y, Nelken I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: Effects of temporal and spectral context. J Neurosci 22:8619–8632.

Bartlett EL, Wang X (2005) Long-lasting modulation by stimulus context in primate auditory cortex. J Neurophysiol 94:83–104.

Blake DT, Merzenich MM (2002) Changes of AI receptive fields with sound density. J Neurophysiol 88: 3409–3420.

Borst A, Flanagin VL, Sompolinsky H (2005) Adaptation without parameter change: Dynamic gain control in motion detection. PNAS 102:6172–6176.

Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation. J Am Stat Assoc 80:580–598.

Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. Neuron 26:695–702.

Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77:923–943.

Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. J Neurophysiol 73:1876–1891.

Christianson GB, Sahani M, Linden J (in press) The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. J Neurosci.

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8:1684–1689.

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. Science 280: 1439–1443.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85:1220–1234.

Effron B, Tibshirani RJ (1993) An introduction to the bootstrap. New York: Chapman & Hall.

Fishbach A, Nelken I, Yeshurun Y (2001) Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transients. J Neurophysiol 85:2303–2323.

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.

Harshman RA, Lundy ME (1994) PARAFAC: Parallel factor analysis. Comput Stat Data Anal 18:39–72.

Hastie TJ, Tibshirani RJ (1999) Generalized additive models. Monographs on Statistics and Applied Probability 43. Chapman & Hall/CRC.

Hunter IW, Korenberg MJ (1986) The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. Biol Cybern 55:135–144.

Jesteadt W, Bacon SP, Lehman JR (1982) Forward masking as a function of frequency, masker level and signal delay. J Acoust Soc Am 71:950–962.

Kowalski N, Depireux DA, Shamma SA (1996a) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76:3503–3523.

Kowalski N, Depireux DA, Shamma SA (1996b) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. J Neurophysiol 76:3524–3534.

Lewicki MS (1994) Bayesian modeling and classification of neural signals. Neural Comput 6:1005–1030.

Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. J Neurophysiol 90:2660–2675.

Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. J Neurosci 24:1089–1100.

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems. New York: Plenum Press.

McKenna TM, Weinberger NM, Diamond DM (1989) Responses of single auditory cortical neurons to tone sequences. Brain Res 481:142–153.

Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87:516–527.

Moore BCJ (1997) An introduction to the psychology of hearing. Academic Press.

Moore BJC (1980) Mechanism and frequency distribution of two-tone suppression in forward masking. J Acoust Soc Am 68:814–824.

Narendra KS, Gallman PG (1966) An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Trans Automat Control AC-11:546–550.

Paatero P (1999) The multilinear engine: a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. J Comp Graph Stats 8:854–888.

Phillips DP. The neural coding of simple and complex sounds in the auditory cortex. In: Sensory Processing in the Mammalian Brain (Lund JS, ed) pp 172–207. Oxford University Press (1989).

Phillips DP, Irvine DRF (1981) Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: Frequency tuning and responses to intensity. J Neurophysiol 45:48–58.

Rotman Y, Bar-Yosef O, Nelken I (2001) Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. Hear Res 152:110–127.

Sahani M, Linden JF. Evidence optimization techniques for estimating stimulus-response functions. In: Advances in Neural Information Processing Systems (Becker S, Thrun S, Obermayer K, eds) volume 15 pp 301–308. Cambridge, MA: MIT Press (2003a). Available online via http://books.nips.cc/.

Sahani M, Linden JF. How linear are auditory cortical responses? In: Advances in Neural Information Processing Systems (Becker S, Thrun S, Obermayer K, eds) volume 15 pp 109–116. Cambridge, MA: MIT Press (2003b). Available online via http://books.nips.cc/.

Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA (2007) Temporal symmetry in primary auditory cortex: Implications for cortical connectivity. Neural Comput 19:583–638.

Smith PH, Populin LC (2001) Fundamental differences between the thalamocortical recipient layers of the cat auditory and visual cortices. J Comp Neurol 436:508–519.

Strang G (1988) Linear algebra and its applications. Harcourt Brace Jovanovich.

Sutter ML (2000) Shapes and level tolerances of frequency tuning curves in primary auditory cortex: Quantitative measures and population codes. J Neurophysiol 84:1012–1025.

Sutter ML, Schreiner CE (1991) Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. J Neurophysiol 65:1207–1226.

Tenenbaum JB, Freeman WT (2000) Separating style and content with bilinear models. Neural Comput 12: 1247–1283.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315–233.

Tomita M, Eggermont JJ (2005) Cross-correlation and joint spectro-temporal receptive field properties in auditory cortex. J Neurophysiol 93:378–392.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6(4):391–398.

Valentine PA, Eggermont JJ (2004) Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. Hear Res 196:119–133.

Vasilescu MAO, Terzopoulos D (2005) Multilinear independent components analysis. Proc IEEE Computer Vision and Pattern Recognition Conf 1:547 – 553.

Wehr M, Zador AM (2005) Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47: 437–445.

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

Young ED, Calhoun BM (2005) Nonlinear modeling of auditory-nerve rate responses to wideband stimuli. J Neurophysiol 94:4441–4454.