

Published in final edited form as:

J Neurosci. 2012 August 15; 32(33): 11271–11284. doi:10.1523/JNEUROSCI.1715-12.2012.

Spectrotemporal Contrast Kernels for Neurons in Primary Auditory Cortex

Neil C. Rabinowitz, Ben D.B. Willmore, Jan W.H. Schnupp, and Andrew J. King

Department of Physiology, Anatomy and Genetics, University of Oxford

Abstract

Auditory neurons are often described in terms of their spectrotemporal receptive fields (STRFs). These map the relationship between features of the sound spectrogram and neurons' firing rates. Recently we showed that neurons in the primary fields of the ferret auditory cortex are also subject to gain control: when sounds undergo smaller fluctuations in their level over time, the neurons become more sensitive to small level changes (Rabinowitz et al., 2011). Just as STRFs measure the spectrotemporal features of a sound that lead to changes in neurons' firing rates, in this study we sought to estimate the spectrotemporal regions in which sound statistics lead to changes in neurons' gain. We designed a set of stimuli with complex contrast profiles to characterize these regions. This allowed us to estimate cortical neurons' STRFs alongside a set of spectrotemporal contrast kernels. We find that these two sets of integration windows match up: the extent to which a stimulus feature causes a neuron's firing rate to change is strongly correlated with the extent to which that feature's contrast modulates the neuron's gain. Adding contrast kernels to STRF models also yields considerable improvements in the ability to capture and predict how auditory cortical neurons respond to statistically complex sounds.

Introduction

One of the central questions that we ask about sensory neurons is what stimulus features they encode in their spike trains. When characterizing neurons throughout the auditory pathway, modellers and electrophysiologists have long employed the spectrotemporal receptive field (STRF) to answer this question (Aertsen et al., 1980; Aertsen and Johannesma, 1981; deCharms et al., 1998; Theunissen et al., 2000; Klein et al., 2000; Miller et al., 2002; Escabí and Schreiner, 2002; Linden et al., 2003; Fritz et al., 2003; Gill et al., 2006; Christianson et al., 2008; Gourévitch et al., 2009; David et al., 2009). The success of STRFs at this task, however, has been somewhat limited (Sahani and Linden, 2003; Machens et al., 2004), necessitating the development of nonlinear extensions – such as adding input nonlinearities (Ahrens et al., 2008b), output nonlinearities (Atencio et al., 2008; Rabinowitz et al., 2011), feedback kernels (Calabrese et al., 2011), simplified second-order interaction terms (Ahrens et al., 2008a), and multiple feature dimensions (Atencio et al., 2008).

One reason for the STRF's limited predictive power is that the encoding of stimulus features by auditory neurons is modulated by stimulus context (Blake and Merzenich, 2002; Valentine and Eggermont, 2004; Ahrens et al., 2008a; Gourévitch et al., 2009). For neurons

Corresponding author: Andrew J.King Department of Physiology, Anatomy and Genetics, University of Oxford Sherrington Building, Oxford OX1 3PT, United Kingdom +44 1865 272 523, ajk@physiol.ox.ac.uk.

Conflicts of interest: None

Commercial Interest: No

in the mammalian primary auditory cortex, the statistics of recent stimulation are a major modulatory influence on the encoding of sound. We recently described a gain control process that is in place by this stage of the auditory pathway (Rabinowitz et al., 2011): neurons in ferret auditory cortex adjust their gain according to the contrast of sound stimulation. When sounds, on average, only change in level by a small amount over time, the neurons scale up their sensitivity to the small fluctuations in sound level. Other authors have observed similar compensatory effects when changing stimulus statistics, from the auditory periphery (Joris and Yin, 1992) to the midbrain (Rees and Møller, 1983; Kvale and Schreiner, 2004; Dean et al., 2005; Nelson and Carney, 2007; Dahmen et al., 2010) and the higher auditory pathway (Nagel and Doupe, 2006; Malone et al., 2010).

Just as STRFs estimate which features of a spectrotemporally complex stimulus drive a neuron to spike, we might ask a similar question of gain changes. What features of a spectrotemporally complex stimulus drive a neuron to change its gain? Our previous work demonstrated, at a coarse, population level, that gain changes are predominantly driven by contrast in sound frequency bands that are local to cortical neurons' best frequencies. However, we do not know how this dependency operates on a neuron-by-neuron basis, or what its relationship is to individual neurons' STRFs.

To answer these questions, we recorded from neurons in the primary auditory fields of the anesthetized ferret, while presenting a set of stimuli with complex patterns of contrast. For each neuron, we determined the spectrotemporal window within which sound contrast informs that neuron's gain. We did this by extending the notion of the STRF, and estimating a set of "gain receptive fields", i.e. spectrotemporal kernels for stimulus contrast. This class of *contrast kernel* models extends the linear-nonlinear (LN) framework of models by capturing the modulation of neurons' input/output functions by patterns of stimulus statistics.

Several possibilities could have arisen. The gain of neurons may be a function of the sound statistics in a broad or a narrow set of frequency bands, and may depend only on the statistics within the short time windows of STRF or on those over longer periods. Our results reveal the relationship between the range of stimulus features that auditory cortical neurons encode, and the range of stimulus statistics which modulate this encoding.

Materials and Methods

Animals

All animal procedures were approved by the local ethical review committee and performed under license from the UK Home Office. Full surgical procedures are provided in Bizley et al. (2010). Briefly, three female adult pigmented ferrets were chosen for electrophysiological recordings under ketamine (5 mg/kg/hr) and medetomidine (0.022 mg/kg/hr) anesthesia. Bilateral extracellular recordings were made in the two auditory cortices using silicon probe electrodes (Neuronexus Technologies, Ann Arbor, MI) with 16 sites on a single probe, vertically spaced at 50 μm . Spikes were sorted offline using spikemonger, an in-house software package. Stimuli were presented via earphones, as described in Rabinowitz et al. (2011).

Stimuli

The main stimulus used was a variant of the dynamic random chord (DRC) stimuli presented in Rabinowitz et al. (2011), which we define here as *Random Contrast DRCs* (RC-DRCs). As with ordinary DRCs, RC-DRCs comprise a sequence of chords, composed of tones whose levels were drawn from particular distributions. For these RC-DRCs, we used $N_F = 23$ pure tones, with frequencies log-spaced between $f_{low} = 500$ Hz and $f_{high} =$

22.6 kHz at 1/4 octave intervals. The levels of the tones were changed every 25 ms, with 5 ms linear ramps between chords. As in Rabinowitz et al. (2011), the amplitude of each tone was always non-zero.

The major distinguishing feature of RC-DRCs is the organisation of these chords into segments of several seconds' duration. In each segment, the distribution of levels for each of the N_F bands had different parameters. A random subset of N_{high} of the N_F tones had their levels drawn from a high contrast (half-width $w_L = 15$ dB; standard deviation $\sigma_L = 8.7$ dB; contrast $c = 92\%$) uniform level distribution, while the remaining $N_{low} = N_F - N_{high}$ tones had their levels drawn from a low contrast ($w_L = 5$ dB; $\sigma_L = 2.9$ dB; $c = 33\%$) uniform level distribution. Both tone distributions had mean level $\mu_L = 40$ dB SPL; these are shown in Figure 1C. By virtue of the 3 s duration, each segment consisted of a sequence of 120 chords, sufficient for a rough approximation of the output nonlinearity during that contrast condition (as explained below).

In order to explore as large a region of contrast space as possible, between $N_S = 80$ and $N_S = 120$ segments were presented at each electrode penetration. Two types of segment were necessary to establish baselines for gain measurements: one where all tone distributions were low contrast, and one where all were high. Given the importance of these two baseline conditions, 9 of the N_S segments were reserved for each. The remaining segments all had a randomized partition of tones into $N_{high} = 5$ high contrast bands and $N_{low} = 18$ low contrast bands, as described above. Thus, the set of segments provided an ensemble of contrast conditions in an analogous way to how an ordinary DRC would provide an ensemble of tone level conditions.

The segments were packaged into individual RC-DRC sequences, each consisting of 12 segments. The first segment of each sequence was 5 s in duration, so that the first 2 s of each stimulus presentation could be discarded. This was necessary as units often showed transient responses to the onset of each DRC sequence that depended on the duration of silence since the end of the last sequence presentation (typically 1-2 s). From each 3 s segment, the first 0.5 s of data was set aside for the analysis of Temporal Contrast Kernels (TCKs), with the remaining 2.5 s of data used to fit Spectral Contrast Kernels (SCKs). The 38 s sequences were presented 10 times each, randomly interleaved.

Unit selection criteria

Only units which modulated their firing rate in response to the RC-DRCs in a reliable, repeatable manner were included for analysis. This was measured via the noise ratio (NR; see Sahani and Linden (2003) and Rabinowitz et al. (2011)) for the PSTH of each unit:

$$\text{noise ratio} = \frac{\text{noise power}}{\text{signal power}} = \frac{\text{total variance} - \text{explainable variance}}{\text{explainable variance}} \quad (1)$$

The PSTH was binned at 25 ms, with bins offset by between 0 and 25 ms to allow for response latency. The offset was chosen on a unit-by-unit basis to minimize the NR. This same offset was used to bin all PSTHs throughout the study; fixing offsets at 10 ms produced similar results. The maximum admitted NR was 40 (estimated across the whole ensemble of stimuli); units with $NR > 40$, i.e. whose explainable variance was less than ~2.5% of the total variance, were excluded from analysis. Models were evaluated while taking NR into consideration (see below).

Notation

We use the following notation throughout this paper. Each DRC stimulus grid is uniquely identified by a matrix (i.e. 2-tensor), L_{tf} , where each component of the matrix describes the sound pressure level (in dB SPL) of a tone with frequency f at time t . To simplify the notation used for fitting spectrotemporal receptive fields (STRFs) below, we define the 3-tensor L_{tfh} as a time-lagged version of L_{tf} : where h is a history index, and the elements of L_{tfh} are defined as the elements of L_{tf} from h time bins in the past, i.e. $L_{(t-h),f}$. The (trial-averaged) response is denoted as y_t , and any model predictions of this response are denoted \widehat{y}_t . Once the STRF was fitted, it was fixed; the output of the STRF model for a given unit is denoted x_t throughout.

As with the tone levels, the contrast profile of the stimulus is denoted by σ_{tf} ; this matrix (or 2-tensor) defines the contrast of the level distribution for the tone at frequency f and time t . Since only two distributions were presented, we define $\sigma = 0$ for the low contrast distribution (yellow in Figure 1B-D) and $\sigma = 1$ for the high contrast distribution (red in Figure 1B-D). Similarly, the recent history of contrast is denoted by σ_{tfh} , whose elements are defined by the time lagged contrast profile, as $\sigma_{tfh} = \sigma_{(t-h),f}$.

Model structure: STRFs and the LN model

The models developed in this paper begin with spectrotemporal receptive fields (STRFs). These were estimated by correlating the stimulus history, L_{tfh} , with the spike peri-stimulus time histogram (PSTH), y_t , at a 25 ms resolution. This involved fitting the general model:

$$\widehat{y}_t = \sum_{f,h} L_{tfh} \cdot k_{fh} \quad (2)$$

STRFs that are separable in frequency, f , and time history, h , often provide better fits than fully in-separable STRFs (Linden et al., 2003; Simon et al., 2007; Ahrens et al., 2008a; Rabinowitz et al., 2011). This was generally the case for this dataset as well. Thus, we assume $k_{fh} = k_f \otimes k_h$ where \otimes is the outer product. This is illustrated in Figure 2B.

We also fitted the majority of the models presented here using inseparable kernels as the first stage of the linear-nonlinear and contrast kernel models. Prediction scores for these models (evaluated using Equation (23) below) were typically 2 – 5 percentage points lower than the corresponding models fitted using separable kernels. Nevertheless, the general trends as presented in this paper were the same (data not shown).

For comparison with the contrast kernel models developed below, the linear STRF was refined by fitting a static linear-nonlinear (LN) model to units' responses (Chichilnisky, 2001; Simoncelli et al., 2004). This involved passing the output of the linear model, x_t , through a static (i.e. memory-less), nonlinear function F , such that $\widehat{y}_t = F[x_t]$. As per Rabinowitz et al. (2011), a logistic curve (sigmoid) was fitted to the data via gradient descent:

$$\widehat{y}_t = a + \frac{b}{1 + \exp\left(-\left(\frac{x_t - c}{d}\right)\right)} \quad (3)$$

The parameters a through d are illustrated in Figure 2C. They can be interpreted as follows: a , as the minimum firing rate; b , as the output dynamic range; c , as the stimulus inflection point; and d , as the (inverse) gain.

Model structure: contrast kernels

To consider how the ongoing contrast profile of the stimulus affects the coding of a cortical neuron, we extended the static linear-nonlinear model above by rendering each of the four parameters, a through d , dependent on the recent history of contrast:

$$\widehat{y}_t = a[\sigma_{tfh}] + \frac{b[\sigma_{tfh}]}{1 + \exp\left(-\left(\frac{x_t - c[\sigma_{tfh}]}{d[\sigma_{tfh}]}\right)\right)} \quad (4)$$

There is considerable freedom in Equation (4) to specify the form of the functions $a[\sigma_{tfh}]$ through $d[\sigma_{tfh}]$. The simplest assumption, which we consider here, is that these are linear functions of σ_{tfh} . This is motivated by three factors: symmetry with the STRF; a linearization of the results in Rabinowitz et al. (2011); and simplicity. The full model takes the form:

$$\widehat{y}_t = a_t + \frac{b_t}{1 + e^{-(x_t - c_t)/d_t}} \quad (5)$$

$$a_t = a_0 + a_1 \cdot \left[\sum_{f, h} \kappa_{fh}^{(a)} \cdot \sigma_{tfh} \right] \quad (6)$$

$$b_t = b_0 + b_1 \cdot \left[\sum_{f, h} \kappa_{fh}^{(b)} \cdot \sigma_{tfh} \right] \quad (7)$$

$$c_t = c_0 + c_1 \cdot \left[\sum_{f, h} \kappa_{fh}^{(c)} \cdot \sigma_{tfh} \right] \quad (8)$$

$$d_t = d_0 + d_1 \cdot \left[\sum_{f, h} \kappa_{fh}^{(d)} \cdot \sigma_{tfh} \right] \quad (9)$$

For brevity of notation, we use the generic parameter θ to denote each of the four nonlinearity parameters, a , b , c , and d . Thus Equations (6) – (9) can be written:

$$\theta_t = \theta_0 + \theta_1 \cdot \left[\sum_{f, h} \kappa_{fh}^{(\theta)} \cdot \sigma_{tfh} \right] \quad (10)$$

Since the profile of recent contrast, σ_{tfh} , varies with time, so each parameter $\theta \in \{a, b, c, d\}$ of the output nonlinearity varies with time. These changes are mediated via a weighted sum of the contrasts in different frequency bands, provided by the term $\kappa_{fh}^{(\theta)}$, which we refer to as the *spectrotemporal contrast kernel* (STCK) for the parameter θ . The form of this model is illustrated in Figure 2A.

As with the STRF, the number of parameters of the contrast kernels can be dramatically reduced by assuming that they are separable in frequency and time history. Thus we constrained $\kappa_{fh}^{(\theta)} = \kappa_f^{(\theta)} \otimes \kappa_h^{(\theta)}$, such that the spectrotemporal contrast kernel could be

decomposed into the outer product of a *spectral contrast kernel* (SCK) and a *temporal contrast kernel* (TCK). We took further advantage of this property by fitting the SCK and TCK separately.

The full model of Equations (5) – (9) has a large number of parameters, numbering $4 \times (N_F + N_H + 1)$ parameters (there being redundancy between $\kappa_{fh}^{(\theta)}$ and θ_1 , as discussed below), in addition to the $N_F + N_H$ parameters of the separable STRF. We made several assumptions to reduce the number of parameters. First, not all of the nonlinearity parameters $\{a, b, c, d\}$ need to be contrast-dependent. For such parameters θ , we set $\theta_1 = 0$, such that $\theta_t = \theta_0$. Next, it is possible that changes to some of these parameters are the result of the same physiological process. This would allow us to assume a shared contrast kernel between pairs of parameters θ and θ' , with $\kappa_{fh}^{(\theta\theta')} \equiv \kappa_{fh}^{(\theta)} = \kappa_{fh}^{(\theta')}$.

For brevity, we assign the following notation to individual models. The full model, wherein all nonlinearity parameters have separate contrast kernels, is denoted as the *a/b/c/d*-model. When a nonlinearity parameter is assumed to be contrast-independent, we omit the corresponding letter from the name. Thus, *b* does not change with contrast in the *a/c/d*-model. Finally, we concatenate letters when they share the same contrast kernel. Thus, in the *a/cd*-model, $\kappa_{fh}^{(a)} \neq \kappa_{fh}^{(c)} = \kappa_{fh}^{(d)} \equiv \kappa_{fh}^{(cd)}$.

One special case is worth particular mention. The results of Rabinowitz et al. (2011) demonstrate that the primary effects of changing contrast lie in changes to the gain, via the parameter *d*, with some correlated changes in threshold, via the parameter *c*. These effects can be most simply captured by the *cd*-model, wherein the nonlinearity parameters *a* and *b* are contrast-independent, while *c* and *d* share a single contrast kernel. The *cd*-model thus takes the form:

$$\begin{aligned} \widehat{y}_t &= a + \frac{b}{1 + e^{-(x_t - c_t)/d_t}} \\ c_t &= c_0 + c_1 \cdot \sum_{f, h} \left[\kappa_{fh}^{(cd)} \cdot \sigma_{tfh} \right] \\ d_t &= d_0 + d_1 \cdot \sum_{f, h} \left[\kappa_{fh}^{(cd)} \cdot \sigma_{tfh} \right] \end{aligned} \quad (11)$$

Fitting procedures

With the assumption that contrast kernels could be separated into spectral (SCK) and temporal (TCK) components, we first fitted SCKs. We limited the TCKs to cover 500 ms of history; as a consequence of both this and the segmented structure of the RC-DRCs, the values of the parameters a_t through d_t would be constant from 500 ms after each segment transition until the next segment transition. We thus fitted SCKs by using only the last 2.5 s of data from each segment, and the following set of equations:

$$\widehat{y}_t = a_t + \frac{b_t}{1 + e^{-(x_t - c_t)/d_t}} \quad (12)$$

$$\theta_t = \theta_0 + \theta_1 \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot \sigma_{tf} \right] \quad (13)$$

In order to reduce the time taken to fit models, we took further advantage of the segmented nature of the RC-DRCs. Since σ_{rf} does not change with time within a segment k , the set of contrast values can be summarized as a matrix s_{kf} capturing the contrast in segment k of frequency band f . Rather than fitting the parameters directly to the entire, trial-averaged training dataset (with $N_T \approx 8000$), the set of (STRF-weighted-)stimulus/response pairs (x_t, y_t) within each segment was divided equally into 20 bins along the x axis (Chichilnisky, 2001; Simoncelli et al., 2004). This reduced the size of the dataset fivefold, and enabled us to confirm that the sigmoid parameterization was appropriate (e.g. Fig. 5, middle column). For bin j in segment k , we denote the bin centre as \bar{x}_{jk} , and the mean firing rate \bar{y}_{jk} . This resultant model was considerably more efficient to fit:

$$\widehat{y}_{jk} = a_k + \frac{b_k}{1 + e^{-(\bar{x}_{jk} - c_k)/d_k}} \quad (14)$$

$$\theta_k = \theta_0 + \theta_1 \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot s_{kf} \right] \quad (15)$$

This reduction in the size of the dataset was necessary for the bootstrapping and Markov Chain Monte Carlo (MCMC) analyses (explained below). On a subset of units we confirmed the validity of this approximation by comparison with fits to Equations (12)–(13). These produced near identical results.

Equations (13) and (15) each contain a redundancy between θ_1 and $\kappa_f^{(\theta)}$. We therefore constrained each SCK to sum to unity, i.e. $\sum_f \kappa_f^{(\theta)} = 1$. For the purposes of including priors (see below), we defined $\lambda_f^{(\theta)}$ as the *unnormalized SCK* for θ , such that:

$$\kappa_f^{(\theta)} = \frac{\lambda_f^{(\theta)}}{\sum_{f'} \lambda_{f'}^{(\theta)}} \quad (16)$$

To fit TCKs for each unit, we returned to the first 0.5 s of data that followed each segment transition. We fixed the values of θ_0 and θ_1 for each θ , together with the SCKs, $\kappa_f^{(\theta)}$, that had already been fitted for each unit. To ensure consistency with Equation (13), we also constrained each TCK to sum to unity, i.e. $\sum_h \kappa_h^{(\theta)} = 1$, by defining $\lambda_h^{(\theta)}$ as the *unnormalized TCK* for θ , via:

$$\kappa_h^{(\theta)} = \frac{\lambda_h^{(\theta)}}{\sum_{h'} \lambda_{h'}^{(\theta)}} \quad (17)$$

As TCKs could only be fitted to the periods immediately after each segment transition, there were limited data available to fit the TCK parameters. One consequence of this was that allowing the coefficients of $\lambda_h^{(\theta)}$ to take on any value (subject to a Gaussian prior) resulted in considerable overfitting (see the \mathfrak{R} performance in Figure 7G). Thus TCKs were fitted with the constraint that all coefficients be positive.

The dataset for each unit was subdivided randomly into training (90%) and prediction (10%) subsets. All parameter fitting took place on the training dataset. Separable STRF models were first fitted to the whole training dataset using maximum likelihood, ignoring the

segmented structure of the RC-DRCs. STRFs were fixed thereafter. Next, Maximum A Posteriori (MAP) estimates of the nonlinearity and contrast kernels were estimated all together, using gradient descent. For each model, the log posterior probability was calculated, as well as its derivatives with respect to all the parameters. Minimization of the negative log posterior was performed using the L-BFGS-B algorithm (Zhu et al., 1997), via SciPy (Jones et al., 2001). This assumed that x_b i.e. the output of the linear STRF model, was observed. The log likelihoods (and the log posteriors) were only convex with respect to some parameters; the gradient descent algorithm thus needed to be initialized at a number of different initial conditions to reduce the chances of settling in local minima. Forty different starting locations were chosen by random draws from the prior distributions over the parameters, with an additional initialization condition at the mean of the priors (see below). Generally, at least half of these repeats converged to the same (best) fixed point.

In principle, we could have merged the STRF fitting with the nonlinearity/contrast kernel fitting, and minimized a single objective function. However, optimizing all parameters proved computationally impractical. Alternatively, we could have iterated between optimizing the nonlinearity/contrast kernel parameters (with the STRF fixed), and optimizing the STRF (with the other parameters fixed). However, we observed that this iterative procedure typically made little to no difference to prediction scores, and STRFs did not noticeably change over successive iterations. Since the focus here is not on the STRF, but on the contrast-dependent changes in output nonlinearities, no successive refinements to STRFs were pursued beyond the initial fit.

A major goal of this work was to characterize the contrast kernels for cortical neurons. This involves estimating not only the best parameter values for $\kappa_f^{(\theta)}$, but also their error bounds. In addition to the MAP estimates, which pinpoint the mode of the posterior parameter distributions, we approximated the shape of these posterior distributions by sampling from them using Markov Chain Monte Carlo (MCMC) methods. MCMC models were constructed in Python, using the PyMC package (Patil et al., 2010). Chains were initialized at the MAP parameter values, and advanced using a Metropolis-Hastings step method. A barrage of diagnostics, including traceplots, Geweke's diagnostic (Geweke, 1992), and autocorrelation analyses, was used to assess convergence and mixing. From these diagnostics, we found that minimum chain lengths of 120000 samples, with a 20000-sample burn-in and 20x thinning were sufficient for a reasonable characterisation of the posteriors. As always for MCMC methods, longer and parallel chains would improve the representation of the posteriors; nevertheless, the observed results satisfied the above diagnostics, and are therefore used here to provide an approximate measure of the error bounds on the contrast kernels. When illustrated in figures, and for computing statistics, these error bounds are summarized in terms of credible intervals, a Bayesian analogue of confidence intervals (Carlin and Louis, 2009).

Priors on nonlinearity parameters

Priors were chosen for simplicity of form, rather than analytic tractability. From Equations (6)–(9), we see that each parameter $\theta \in \{a, b, c, d\}$ is the sum of a contrast-independent term, θ_0 , and a contrast-dependent term, weighted by θ_1 . Rather than placing priors directly on these terms, it was more convenient to reparameterize the model as follows.

The contrast in segment k and frequency f , s_{kf} , could only take on binary values. In the all-low-contrast segment, when $s_{kf} = 0 \forall f$, it follows from Equation (15) that $\theta_k = \theta_0$. In the all-high-contrast segment, when $s_{kf} = 1 \forall f$, we use the fact that the contrast kernel, $\kappa_f^{(\theta)}$, is

normalized (from Equation (16)) to find that $\theta_k = \theta_0 + \theta_1$. Defining these two values as θ_{low} and θ_{high} respectively, we can rewrite Equation (15) as:

$$\theta_k = \theta_{low} + (\theta_{high} - \theta_{low}) \cdot \left[\sum_f \kappa_f^{(\theta)} \cdot s_{kf} \right] \quad (18)$$

Thus the parameters $\{a_{low}, b_{low}, c_{low}, d_{low}\}$ describe the output nonlinearity in the all-low-contrast segment, and the parameters $\{a_{high}, b_{high}, c_{high}, d_{high}\}$ describe the output nonlinearity in the all-high-contrast segment. In each segment, the value of θ_k typically lies between θ_{low} and θ_{high} , depending on the projection of s_{kf} onto $\kappa_f^{(\theta)}$ (although θ_k can take on more extreme values when some of the coefficients of $\kappa_f^{(\theta)}$ are negative).

For each θ , identical priors were placed on each of θ_{low} and θ_{high} . The respective priors were primarily chosen to satisfy three purposes: to enforce a set of hard constraints, namely that a , b and d all be positive; to apply some regularization, i.e. to ensure that b , c and/or d did not grow excessively large; and to provide a suitable set of initial conditions for MAP fitting. As a result, the priors were relatively broad, with data-driven hyperparameters.

For each unit, we defined a set of intermediate statistics on the binned stimulus-response data, $(\bar{x}_{jk}, \bar{y}_{jk})$:

$$\begin{aligned} y_{min} &= \min_{j,k} \bar{y}_{jk} & y_{min} &= \min_{j,k} \bar{x}_{jk} \\ y_{max} &= \max_{j,k} \bar{y}_{jk} & x_{max} &= \max_{j,k} \bar{x}_{jk} \\ y_{range} &= y_{max} - y_{min} & x_{range} &= x_{max} - x_{min} \\ & & x_{mean} &= \frac{1}{N_j N_k} \sum_{j,k} \bar{x}_{jk} \end{aligned}$$

where N_j is the number of bins (here, 20), and $N_k = N_S$, the number of segments.

In turn, the priors $P(\theta_{low}) = P(\theta_{high})$ were defined via:

$$a_{low}, a_{high} \sim \text{Exp}(y_{min} + 0.05 \times y_{range}) \quad (19)$$

$$b_{low}, b_{high} \sim \text{Exp}(2 \times y_{range}) \quad (20)$$

$$c_{low}, c_{high} \sim \mathcal{N}(x_{mean}, x_{range}^2) \quad (21)$$

$$d_{low}, d_{high} \sim \text{Exp}(x_{range}/20) \quad (22)$$

where exponential distributions are given in terms of their scale parameters, β .

The L-BFGS-B algorithm used to minimize the negative log posterior allows the explicit specification of parameter boundary values; for the exponentially distributed variables, a lower bound of 1×10^{-15} was provided.

Finally, for those models where only a subset of the nonlinearity parameters were contrast-dependent, $\theta_1 = 0$ was enforced, such that $\theta_{high} = \theta_{low}$.

Priors on contrast kernels

As discussed in the results, three different approaches to the values of SCKs were taken. In the first approach, the kernels were allowed to take any real value; in the second approach, they were constrained to be positive; in the third approach, they were fixed at particular values.

Priors were placed on the coefficients of the unnormalized contrast kernels, $\lambda_f^{(\theta)}$. As these were normalized via Equation (16) to give $\kappa_f^{(\theta)}$, the scale of the respective priors was not important.

When real-valued kernels were used, the prior on $\lambda^{(\theta)}$ was chosen to be a spherical Gaussian, with $\lambda_f^{(\theta)} \sim \mathcal{N}(0, 0.1^2)$. When positive-valued kernels were used, the prior on each coefficient was chosen as $\lambda_f^{(\theta)} \sim \text{Exp}(0.1)$. In this latter situation, as for the positive nonlinearity parameters, the L-BFGS-B algorithm bounded each coefficient below at a value of 1×10^{-15} .

When real-valued contrast kernels were used, it was possible for the denominator of Equation (16) to approach zero, giving untenable values of $\kappa_f^{(\theta)}$. As a result, the minimization algorithm occasionally yielded zero-division errors. When this occurred, the algorithm was reset with a new initial value. This discontinuity also meant that the algorithm was more likely to get stuck in local minima of the negative log posterior, requiring a larger number of repeated fittings from random initial conditions.

Model success

In order to compare different models for the firing rate behaviour of auditory cortical neurons, we made use of the strategy developed by Sahani and Linden (2003). For each unit, the amount of its total response variance that can be explained is bounded by the signal power (SP). Model success should therefore be measured as the percentage of signal power explained (%SPE). This is the percentage reduction in the SP from fitting the model, and is equivalent to:

$$\%SPE = \frac{\text{Var}(y_t) - \text{Var}(y_t - \hat{y}_t)}{SP} \times 100\% \quad (23)$$

By subdividing the data for each unit into a training and prediction dataset, one can obtain two values for %SPE for that unit. The %SPE from the training data is inflated as a result of overfitting to the noise in the training data. Thus %SPE_{training} provides an upper bound for the model performance. The %SPE from the prediction data is expected to be lower, since it tests the generalisability of the model to new data. Thus %SPE_{prediction} provides a lower bound for the model performance.

Since these two measures diverge as a function of noise ratio, a suitable method for measuring the model's predictive power from the population data is to extrapolate from the two sets of estimates of model success above to those for a hypothetical zero-noise neuron (Sahani and Linden, 2003; Ahrens et al., 2008a). The resultant upper and lower estimates bound the *true predictive power* of the model, i.e. that which would be obtained in the limit of zero noise. Bounds of model prediction power reported here come from linear

extrapolations to the zero case. Where only a single value is cited (as in the figures), this is the lower bound.

To correct against sampling biases, we cross-validated the results across ten different partitions of the data. The %*SPE* values reported here are medians across these ten partitions. The same set of partitions were used for fitting all models to the same unit.

Results

Our primary objective was to determine the spectrotemporal window within which changes in stimulus contrast inform changes in neuronal gain. To do so, we designed a set of stimuli, known as Random Contrast Dynamic Random Chord (RC-DRC) sequences (Fig. 1). This provided an ensemble of stimulation conditions, each with a different profile of contrast statistics.

We recorded from 168 units in the primary auditory cortex (A1) and anterior auditory field (AAF) of three anesthetized ferrets, while presenting RC-DRCs. These areas were identified on the basis of their location on the middle ectosylvian gyrus and the tonotopic organization, which is organized approximately dorso-ventrally across the gyrus (Nelken et al., 2004; Bizley et al., 2005). Amongst this set of units, we identified 77 units which responded reliably to the RC-DRCs, as measured via a maximum noise level criterion (see Materials and Methods).

Spectral contrast kernels

We constructed a class of models to analyse units' responses to the RC-DRCs (Equations (5)–(9); Fig. 2). These build upon linear-nonlinear (LN) models, which have previously been used to characterize the relationship between stimuli and neuronal responses (Chichilnisky, 2001; Simoncelli et al., 2004). As in a standard LN approach, we modelled units' responses as a two stage process: first, a reduction of the dimensionality of stimulus space, by filtering the ongoing (log) spectrogram through a spectrotemporal receptive field (STRF); second, a nonlinear transformation stage, by passing the filtered stimulus through a static output nonlinearity. Our models expanded upon this schema by allowing the parameters of the output nonlinearity — and therefore its shape — to change over time. In particular, we enabled these parameters to change as a function of stimulus statistics. Because the RC-DRC stimuli were constructed by defining a matrix of contrast statistics, which varied over frequency and time, we modelled the changes to the nonlinearity parameters via a set of *spectrotemporal contrast kernels* (STCKs). Each STCK filters the ongoing contrast profile of the sound, σ_{tf} , in the same way that the STRF filters the ongoing spectrogram, L_{tf} .

The most general model of this scheme has a large number of parameters. We therefore began by making a few key simplifications. First, we assumed that spectrotemporal contrast kernels could be separated into a spectral component and a temporal component, in the same way that it is often reasonable to make separable approximations to cortical STRFs (see above). Thus, we first fitted spectral contrast kernels (SCKs), and later temporal contrast kernels (TCKs).

Our second simplification was to consider, for each parameter of the output nonlinearity, whether that parameter showed evidence of being dependent on stimulus contrast. This was motivated by our previous results (Rabinowitz et al., 2011) which showed that changing the global stimulus contrast primarily produced changes in gain (here, the d parameter of the nonlinearity) and the stimulus inflection point (here, the c parameter). Finally, further reductions in the parameter load could be made by sharing contrast kernels between multiple parameters.

To assess the validity and utility of such simplifications, we fitted a range of SCK models to the responses of the cortical units. For each model and unit, we measured the fit quality, together with its ability to predict responses outside of the training dataset. These were quantified as the percentage of stimulus-locked response variance that the model explained in each of the two datasets (Equation (23)). Previous authors have demonstrated that the measured values of such quantities depend on the trial-to-trial reliability of units' stimulus-evoked spiking patterns: for less reliable (i.e. noisier) units, fitted models are more likely to capture noise in the training dataset, and therefore make poorer predictions (Sahani and Linden, 2003). We thus followed Sahani and Linden's lead, and assessed model performance across the population of cortical units by extrapolating from the set of scores to an idealized, zero-noise unit. This produced two estimates of the model's predictive power: an upper bound, from the model's fit quality on the training sets, and a lower bound, from the model's ability to predict outside the training sets (Ahrens et al., 2008a). This process is illustrated in Figure 3A for the lower bounds.

As a baseline, we fitted simple (separable) STRF and LN models to each unit. These models were fitted to data that were pooled across all segments of the RC-DRCs, and therefore did not take into account changes in contrast from segment to segment. The predictive power of the STRF model was 42.4–43.4%, while the predictive power of the LN model was 60.2–62.2%. Adding an output nonlinearity considerably improves model performance.

Including a full set of independent SCKs for each nonlinearity parameter also improved model predictions (Fig. 3). This *a/b/c/d*-model (see Materials and Methods for naming conventions) had a prediction score of 62.7 – 70.1%. However, we found that we could substantially further improve the models' predictive performance by adding constraints to reduce the degree of overfitting. First, we found that the parameters *a* and *b* did not generally change with contrast. Fixing these to be contrast-independent (i.e. fitting the *c/d*-model) yielded better prediction performance of 65.6 – 70.1%. In turn, the spectral contrast kernel for the *c* parameter, $\kappa_f^{(c)}$, and the spectral contrast kernel for the *d* parameter, $\kappa_f^{(d)}$, were generally highly correlated with each other (median correlation coefficient of $r(k_f^{(c)}, k_f^{(d)}) = 0.89$). We therefore constrained these two contrast kernels to be identical (the *cd*-model). This outperformed the other SCK models, with a prediction score of 66.2 – 70.1%. On a unit-by-unit basis, the *cd*-model outperformed the standard LN model for 62/77 units; this improvement was significant for 48 of these units (Wilcoxon signed-rank test on $N = 40$ cross-validated scores, $p < 0.01$).

According to the *cd*-model, 72/77 units decreased their gain as contrast increased. The extent of gain changes can be quantified as the ratio $G_d = d_{high}/d_{low}$ (see Equation (18)), which measures the proportional dilation of the output nonlinearity along the *x*-axis as a result of switching from the all-high-contrast condition to the all-low-contrast condition. A histogram of G_d values for the population of units is shown in Figure 4A. The median G_d was 1.92, which is in good agreement with our previous observations (Rabinowitz et al., 2011). As expected, units with larger G_d tended to experience the greatest improvements in model prediction by including the spectral contrast kernel (Fig. 4B; Spearman correlation of 0.40; $p < 0.001$).

In summary, the most parsimonious model for capturing contrast-dependent changes to units' firing behaviour is the *cd*-model of Equation (11). As contrast is varied, auditory cortical neurons' output nonlinearities undergo a slope change and a horizontal shift. These changes can be described as a linear function of the spectral profile of contrast. In the sections which follow, we concentrate exclusively on the *cd*-model.

The shape of SCKs

We next asked what the spectral contrast kernels looked like. Examples of SCKs fitted to cortical units' responses are shown in Figure 5. The most striking aspect of the SCKs is their similarity in shape to the STRF frequency kernels, k_f . For the frequencies in the excitatory component of the receptive field of these units, the weights of the SCK, $\kappa_f^{(cd)}$, match almost precisely the weights of k_f . As the best frequency (BF) and bandwidth of k_f change across these units, so the BF and bandwidth of $\kappa_f^{(cd)}$ change too.

There was generally a good correlation between units' gain SCKs, $\kappa_f^{(cd)}$, and the frequency component of their linear STRF kernels, k_f across units, the median correlation coefficient was $r(\kappa_f^{(cd)}, k_f) = 0.69$.

These kernels thus reveal an important aspect of contrast gain control: the same frequency channels whose level changes *additively* contribute to a cortical unit's firing rate also *divisively* contribute to its gain. In these bands, an increase in tone level *increases* the unit's firing rate, while an increase in the contrast of the tone level distribution of these bands *decreases* the unit's gain. In turn, the relative size of the gain change produced by varying the contrast in a particular band is roughly proportional to the size of the change in firing rate produced by increasing the level of the band.

Contrary to this pattern, we found that when units had strong inhibitory sidebands in their STRF — i.e. when there were coefficients of k_f nearby to the BF, which were negative — the SCKs often had *positive*, rather than negative, coefficients for these same frequencies (e.g. Fig. 5E-G). In these bands, an increase in tone level *decreases* the firing rate of the neuron; however, an increase in the contrast of the tone level distribution of these bands *also decreases* the gain of the neuron.

These qualitative observations capture the major trends we observed. Amongst those units which deviated somewhat from this pattern, some had slightly wider SCKs, and others slightly narrower than the excitatory band of the STRF. In addition, not all of the units with inhibitory sidebands produced significantly non-zero $\kappa_f^{(cd)}$ coefficients at the sidebands (e.g. Fig. 5H). Finally, approximately 20% of units (17/77) produced noisy, random-shaped contrast kernels. Amongst this last group, the SCK models still produced reasonable prediction scores; the results for these units are discussed in more detail below.

Simplifying SCKs

It is clear from the examples of Figure 5 that the most salient features of the gain SCKs are their large, positive coefficients in a localized region of frequency space. As mentioned above, these coefficients are often also positive in the inhibitory sidebands. Increasing the contrast of any of these bands thus yields a decrease in neuronal gain. However, very few $\kappa_f^{(cd)}$ coefficients across the set of models appeared to be genuinely *negative*, such that high contrast in these bands would lead to an *increase* in neuronal gain.

Although 45% of all $\kappa_f^{(cd)}$ coefficients were fitted to negative values, these values were typically small in magnitude. They were also generally not significant: the marginal posteriors on these coefficients rarely had all their weight below zero. In total, according to a 95% credible interval criterion, 7% of $\kappa_f^{(cd)}$ coefficients across all units were significantly negative; according to a 99% credible interval criterion, only 3% of κ_f coefficients were

significantly negative. These values compare with 19% and 13% for significantly positive κ_f coefficients.

Thus, the coefficients of SCKs were rarely negative, and were generally larger in magnitude when the STRF frequency kernel was larger in magnitude. As a result, rather than describing a correlation between the coefficients of units' $\kappa_f^{(cd)}$ and their k_f (as above), there was actually a better correlation between $\kappa_f^{(cd)}$ and the absolute value of the STRF frequency kernel, $|k_f|$ for each unit, with a median correlation coefficient of $r(\kappa_f^{(cd)}, |k_f|) = 0.80$.

There is reason to suspect that none of the coefficients of the gain contrast kernel should be negative. In principle, negative $\kappa_f^{(cd)}$ values indicate frequency bands for which high contrast would cause an *increase* in neuronal gain. This may not be possible under certain mechanistic implementations of contrast gain control. To test this possibility, we enforced the constraint that coefficients of $\kappa_f^{(cd)}$ must be positive. Examples of the resulting kernels are shown in Figure 6A-H. This model provided even better predictions than using real-valued (i.e. unconstrained) $\kappa_f^{(cd)}$, with a prediction score of 67.1–69.9% (Fig. 6I). The constrained-positive *cd*-model was, in total, the best predicting SCK model, and outperformed the standard LN model for 68/77 units. This improvement was significant for 52 of these units ($p < 0.01$). Thus, it is likely that negative values in the unconstrained contrast kernels reflect an overfitting of the parameter values to the small sample of conditions presented.

As mentioned above, a number of units' gain contrast kernels were noisy, with little observable structure when $\kappa_f^{(cd)}$ was unconstrained. For all but five of these units, constraining $\kappa_f^{(cd)} > 0$ yielded contrast kernels which more closely resembled the units' respective STRF frequency kernels. Constraining the coefficients of the gain kernels to be positive therefore reveals an underlying structure to the kernels of noisier units.

A striking feature of the constrained-positive kernels is that across all units that gave reliable responses to RC-DRCs, the correlations between the coefficients of $\kappa_f^{(cd)}$ and $|k_f|$ are even stronger than for the unconstrained models, with a median $r(\kappa_f^{(cd)}, |k_f|) = 0.93$. This suggests that one may be able to approximate the gain contrast kernels simply as $\kappa_f^{(cd)} = |k_f|$. We implemented this as an additional set of models. These showed that when the SCK was fixed in this manner rather than fitted, the model performance was only slightly impaired relative to fitting an SCK, as shown in Figure 6I. Almost identical prediction scores were obtained when we fixed $\kappa_f^{(cd)}$ to be a rectified version of k_f indicating that the contribution from the inhibitory sidebands to the model success was small. Finally, as a control, we also tested three alternative models: one with $\kappa_f = k_f$ i.e. without the absolute value; a second where $\kappa_f = |H(k_f)|$, i.e. as the magnitude of the Hilbert transform of k_f (which produced wider bandwidth SCKs); and a third where we ignored all spectral information and assumed a constant SCK ($\kappa_f = 1/N_f$). In all three cases, the model performed substantially worse. These data are summarised in Figure 6I.

The usefulness of the approximation $\kappa_f^{(cd)} \approx |k_f|$ is important. As this experiment demonstrates, the number of contrast conditions needed to estimate gain contrast kernels is large, making it a time-consuming process. On the other hand, including gain changes leads to substantial increases in model performance. When using the approximation, only two

additional variables beyond the standard out-put nonlinearity need to be estimated (c_1 and d_1), which can be done quickly using only two contrast conditions. Thus, greatly improved models of the responses of auditory cortical neurons can be readily implemented using this approach.

Temporal contrast kernels

Just as the SCKs reveal how units integrate the spectral pattern of stimulus contrast to determine their gain, TCKs reveal how units integrate the recent history of stimulus contrast to the same effect. To map cortical units' TCKs, we fixed their SCKs, and fitted models to the neuronal responses immediately following each segment transition.

Examples of temporal contrast kernels, $\kappa_h^{(cd)}$, are shown in Figure 7A-D. As for the temporal component of these units' STRFs, the units were most sensitive to the contrast in the most recent 50–100 ms of stimulation, and retained a weak dependence on the contrast statistics further back in history.

Including the temporal contrast kernel for gain changes led to an overall improvement in the model predictive power. For the responses during these transition periods, the prediction scores were 43.1% for the STRF model, 59.7% for the linear-nonlinear model, 64.3% when only the SCK was considered, and 67.3% when the full STCK was implemented (Fig. 7F). The STCK outperformed the LN model for 72/77 units, of which 54 were significant (Wilcoxon signed-rank test on $N = 100$ cross-validated scores; $p < 0.01$).

Simplifying TCKs

A secondary goal of this work is to develop simple approximations to contrast kernels that can be applied without requiring the time-consuming exploration of stimulus space attempted here. We therefore considered a number of simplifications to the TCK model. The success of each of these simplifications is summarized in Figure 7G.

We noted that the population mean of the TCKs, shown in Figure 7E, followed an approximately exponential decay, with a time constant of $\tau = 86$ ms. We therefore fitted a simplified, single-parameter TCK model to each unit, $\kappa_h^{(cd)} \propto \exp(-h/\tau_H)$, where τ_H is the time constant (Fig. 7H). The median time constant fitted to the 77 units was $\tau_H = 117$ ms. The model performed well compared with fitting a full TCK, with a score of 67.2%. We also considered whether, for the purposes of parsimony, a single time constant could be used for all units within the population. By fixing τ_H at different values, we found that the most predictive model had $\tau_H = 85$ ms, with a prediction score of 67.1%. There was, however, a reasonably broad range of τ_H values between 80 – 120 ms which gave similarly respectable scores (Fig. 7I).

Finally, in the same way that the SCKs could be approximated, up to a normalisation constant, as the absolute value of the frequency component of the STRF, so too the TCKs could be approximated as the absolute value of the temporal component of the STRF. This produced a prediction score of 67.2%. Thus the absolute value of the STRF provides an excellent approximation for the spectrotemporal contrast kernel of a cortical neuron.

Discussion

The goal of this study was to determine the spectrotemporal windows within which stimulus contrast modulates the gain of auditory cortical neurons. We therefore constructed a stimulus set that provided an ensemble of different contrast conditions (Fig. 1), and investigated how cortical units' response properties changed under these conditions (Fig. 2).

We were able to estimate the relative contributions of the contrast in different frequency bands and different time bins to the gain of individual units, via their spectrotemporal contrast kernels (STCKs).

We found that the spectral components of these kernels (the SCKs) typically place their weight on the same frequency bands that contribute to a neuron's STRF (Fig. 5). Thus, when a neuron's firing rate is *linearly* sensitive to the level variations in a particular band, then it is also *divisively* sensitive to changes in the contrast of that band. Not only are SCKs co-extensive with the frequency component of the STRF, but they are also matched in magnitude: the extent to which a band's contrast contributes to a neuron's gain is approximately proportional to the extent to which that band's level contributes to the neuron's firing rate (Fig. 6I). Neurons with narrow tuning curves are sensitive to contrast in a narrow frequency window, whereas the gain of neurons with broad tuning curves can be influenced by contrast over a similarly broad frequency range. Curiously, the spectral region whose statistics determine gain includes a neuron's inhibitory sidebands: high contrast in the sidebands also reduces neural gain.

The temporal component of these kernels (the TCKs) could be fitted reasonably well by an exponential curve, with a time constant of ~ 85 ms. Similar to the SCKs, the TCKs could also be approximated well as the absolute value of the time component of the corresponding STRF (Fig. 7). Thus, a simple approximation of the gain contrast kernel is $\kappa_{fh}^{(cd)} \approx |k_{fh}|$. In sum, cortical neurons integrate stimulus contrast and level fluctuations over a similar spectrotemporal window, albeit to different effects. This is summarized in Figure 8.

Little contribution to the gain from remote spectral and temporal regions

This study considerably extends a preliminary estimation of contrast kernels presented in Rabinowitz et al. (2011). There, we attempted a coarse, population-level characterization of auditory cortical neurons' SCKs, and found that neurons' gain depended predominantly on the contrast in spectral regions local to units' best frequencies. This is confirmed by the results presented here.

Although our previous study ruled out strong contributions to the gain from frequency bands outside the neurons' STRFs, we did find evidence for weak contributions from these bands, suggesting that gain control in the auditory cortex is, to some extent, dependent on global statistics. In the present study, however, we found that the gain kernels were largely restricted to the frequency bands present in the STRF.

Stimulus design may explain this discrepancy. In Rabinowitz et al. (2011), we categorically divided frequency bands into local and remote groups, in a way that may have underestimated the range of frequency bands which additively contributed to the STRF. Our approach here circumvented this problem by being non-categorical. On the other hand, the subset of contrast space explored here may have been insufficient to reveal the contributions from remote bands, which could be weak (or superadditive) and only detectable as a compound effect. These results therefore bound the magnitude of extra-classical receptive field contributions to neuronal gain.

The match between the domains of SCKs and linear STRFs is consistent with previous findings on forward suppression: in general, the more a sound matches a neuron's preferred stimulus, the more it suppresses subsequent responses (Calford and Semple, 1995; Brosch and Schreiner, 1997; Reale and Brugge, 2000; Zhang et al., 2005; Scholl et al., 2008). The match between TCKs and STRFs, however, initially seems at odds with the long timescales of adaptation previously reported in the auditory cortex (Ulanovsky et al., 2004; Wehr and

Zador, 2005; Asari and Zador, 2009). The contrast kernel models therefore capture only a fast component of this adaptation, much like the rapid luminance and contrast gain control identified in the retina (Enroth-Cugell and Shapley, 1973; Baccus and Meister, 2002). It is possible we did not see slower adaptation components because our DRCs switched contrast rapidly: in the retina, the timescale and parameters of stimulus dynamics directly impact on the timescale of slow contrast adaptation (Wark et al., 2009).

The similarity between the domains of STCKs and STRFs suggest that both phenomena share some common source. However, our results can only partially constrain this mechanism. The shape of STRFs depends on complex interactions between excitation and inhibition (Wallace et al., 1991; Budinger et al., 2000; Winer et al., 2005; Liu et al., 2007; Wu et al., 2008; Moeller et al., 2010). Gain control could therefore be explained by a combination of excitatory and/or inhibitory inputs (Chance et al., 2002; Murphy and Miller, 2003; Katzner et al., 2011), the action of intrinsic currents (Abolafia et al., 2010), or the activation of local layer six neurons with similar tuning, as recently observed in V1 (Olsen et al., 2012). Given the rapidity of the TCKs, our results are unlikely to be fully described by cortical synaptic depression, which appears to operate at longer timescales (Wehr and Zador, 2005). Gain control may have subcortical origins (Malmierca et al., 2009; Anderson et al., 2009), provided these combine in a similar manner to the way they produce cortical STRFs. It may be possible to evaluate the relative likelihood of these mechanisms by comparing STRFs and STCKs under different stimulation conditions since STRFs are known to change under different stimulus contexts (Theunissen et al., 2000; Blake and Merzenich, 2002; Valentine and Eggermont, 2004; Woolley et al., 2005; David et al., 2009; Schneider and Woolley, 2011).

Implications for modelling

The contrast kernel models advanced in this work provide considerably better predictions of neurons' responses compared with STRF and LN models. They capture ~ 20% of the residual variance not explained by the LN model. Since STCKs can be approximated well from the absolute value of the STRF, this model requires only two additional parameters beyond the LN-model (and hence six parameters beyond the STRF). The model presented in Figure 8 thus provides a simple and powerful way of extending existing models for the responses of auditory cortical neurons, capturing these neurons' sensitivity to patterns of stimulus contrast.

The gold standard for models such as these is to be able to predict responses of auditory neurons to natural stimuli (Wu et al., 2006). Studies which have estimated receptive field models using synthetic stimuli have repeatedly found that the models do not generalize well to natural sounds (Theunissen et al., 2000; Rotman et al., 2001; Machens et al., 2004; David et al., 2009). One compelling reason for this is that natural sounds likely engage nonlinear coding mechanisms, which may not be activated within the spaces of synthetic stimuli such as DRCs or ripples (Theunissen et al., 2000; Woolley et al., 2006; David et al., 2009). Furthermore, the linear approximations made during model construction are sensitive to the statistics of the subspace of stimuli explored (Christianson et al., 2008). Since natural scenes vary in their statistics over time, it is likely that including time-varying gain control will improve the predictions of STRF-based models.

One particular difficulty in extending these models to other domains is knowing how to measure stimulus contrast. For the synthetic stimuli we used here, the contrast was specified by design; we therefore used the stimulus parameters as input into the models. For arbitrary sounds, an algorithm for estimating σ_{tth} would need to be specified; provided this algorithm makes broadly consistent measurements of the stimulus parameters that we used here, we anticipate that the benefits of including gain control will be considerable.

Gain control and divisive normalization

One form in which gain control is often cast is that of divisive normalization. In the abstract, this is a gain standardization process by which an initial set of responses — usually the result of information fed forward from earlier brain areas — is rescaled. The scaling factor takes the form of a local response normalizer: the activity of each neuron is divided by the pooled activity over other neurons in a local neighbourhood (Heeger, 1992; Carandini et al., 1997). There is considerable evidence for normalization in a large number of systems, including V1 (Heeger, 1992; Carandini et al., 1997; Rust et al., 2005), extrastriate visual cortex (Miller et al., 1993; Recanzone et al., 1997; Missal et al., 1997; Simoncelli and Heeger, 1998; Britten and Heuer, 1999; Heuer and Britten, 2002; Zoccolan et al., 2005), superior colliculus (Basso and Wurtz, 1997), and the *Drosophila* antennal lobe, which mediates olfaction (Olsen et al., 2010), as well as in multi-sensory integration (Ohshiro et al., 2011).

There remains considerable debate as to what combination of cellular and circuit mechanisms actually mediates divisive normalization in the visual system (Carandini and Heeger, 2011). Nevertheless, it has proved to be a powerful idea for advancing our understanding of the computations actually being performed by a given system. Normalization promotes efficient coding, not only by shifting stimulus representations to use more of the neurons' dynamic range, but also by encouraging decorrelated, higher-entropy representations of natural signals (Ruderman and Bialek, 1994; Olshausen and Field, 1996; Brady and Field, 2000; Fairhall et al., 2001; Schwartz and Simoncelli, 2001). Theoretical work has also argued for a role for normalization in other computations, such as decoding (Deneve et al., 1999; Ringach, 2010) and marginalization (Beck et al., 2011).

If we consider a network implementation of gain control, our result that $\kappa \approx |k|$ demonstrates that auditory cortical neurons have gain pools which share similar spectrotemporal sensitivity profiles. Thus, just as many systems appear to construct representations that are invariant to the normalized statistic, including visual representations in V1 that are contrast-invariant (Albrecht and Hamilton, 1982; Heeger, 1992; Busse et al., 2009; Ringach, 2010), velocity representations in MT that are spatial-pattern-invariant (Heeger et al., 1996; Simoncelli and Heeger, 1998), and odor representations in the antennal lobe that are concentration-invariant (Luo et al., 2010; Olsen et al., 2010), so it appears that the auditory cortex builds representations of sounds that are partially invariant to their spectrotemporally local contrast.

Acknowledgments

This work was supported by the Wellcome Trust through a Principal Research Fellowship to A.J.K. (WT076508AIA) and by Merton College, Oxford through a Domus A three-year studentship to N.C.R. We are grateful to Sandra Tolnai for assistance with data collection. We also would like to thank Fernando Nodal for his helpful contributions to the surgical preparations.

References

- Abolafia JM, Vergara R, Arnold MM, Reig R, Sanchez-Vives MV. Cortical auditory adaptation in the awake rat and the role of potassium currents. *Cerebral Cortex*. 2010; 21:977–990. [PubMed: 20851851]
- Aertsen AMHJ, Johannesma PIM. The spectro-temporal receptive field. *Biological Cybernetics*. 1981; 42:133–143. [PubMed: 7326288]
- Aertsen AMHJ, Johannesma PIM, Hermes DJ. Spectro-temporal receptive fields of auditory neurons in the grassfrog II: analysis of the stimulus-event relation for tonal stimuli. *Biological Cybernetics*. 1980; 38:235–248.

- Ahrens M, Linden J, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *The Journal of Neuroscience*. 2008a; 28:1929–42. [PubMed: 18287509]
- Ahrens MB, Paninski L, Sahani M. Inferring input nonlinearities in neural encoding models. *Network: Computation in Neural Systems*. 2008b; 19:35–67.
- Albrecht DG, Hamilton DB. Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*. 1982; 48:217–237. [PubMed: 7119846]
- Anderson LA, Christianson GB, Linden JF. Stimulus-Specific adaptation occurs in the auditory thalamus. *The Journal of Neuroscience*. 2009; 29:7359–7363. [PubMed: 19494157]
- Asari H, Zador AM. Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. *Journal of Neurophysiology*. 2009:00577.2009.
- Atencio CA, Sharpee TO, Schreiner CE. Cooperative nonlinearities in auditory cortical neurons. *Neuron*. 2008; 58:956–966. [PubMed: 18579084]
- Baccus S, Meister M. Fast and slow contrast adaptation in retinal circuitry. *Neuron*. 2002; 36:909–919. [PubMed: 12467594]
- Basso MA, Wurtz RH. Modulation of neuronal activity by target uncertainty. *Nature*. 1997; 389:66–69. [PubMed: 9288967]
- Beck JM, Latham PE, Pouget A. Marginalization in neural circuits with divisive normalization. *The Journal of Neuroscience*. 2011; 31:15310–15319. [PubMed: 22031877]
- Bizley JK, Nodal FR, Nelken I, King AJ. Functional organization of ferret auditory cortex. *Cerebral Cortex*. 2005; 15:1637–1653. [PubMed: 15703254]
- Bizley JK, Walker KMM, King AJ, Schnupp JWH. Neural ensemble codes for stimulus periodicity in auditory cortex. *The Journal of Neuroscience*. 2010; 30:5078–5091. [PubMed: 20371828]
- Blake DT, Merzenich MM. Changes of AI receptive fields with sound density. *Journal of Neurophysiology*. 2002; 88:3409–3420. [PubMed: 12466457]
- Brady N, Field DJ. Local contrast in natural images: normalisation and coding efficiency. *Perception*. 2000; 29:1041–1055. PMID: 11144818. [PubMed: 11144818]
- Britten KH, Heuer HW. Spatial summation in the receptive fields of MT neurons. *The Journal of Neuroscience*. 1999; 19:5074–5084. [PubMed: 10366640]
- Brosch M, Schreiner CE. Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*. 1997; 77:923–943. [PubMed: 9065859]
- Budinger E, Heil P, Scheich H. Functional organization of auditory cortex in the mongolian gerbil (*Meriones unguiculatus*). III. anatomical subdivisions and corticocortical connections. *European Journal of Neuroscience*. 2000; 12:2425–2451. [PubMed: 10947821]
- Busse L, Wade AR, Carandini M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron*. 2009; 64:931–942. [PubMed: 20064398]
- Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SMN. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE*. 2011; 6:e16104. [PubMed: 21264310]
- Calford MB, Semple MN. Monaural inhibition in cat auditory cortex. *Journal of Neurophysiology*. 1995; 73:1876–1891. [PubMed: 7623087]
- Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*. 2011; 13:51–62.
- Carandini M, Heeger DJ, Movshon JA. Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*. 1997; 17:8621–8644. [PubMed: 9334433]
- Carlin, B.; Louis, T. Bayesian methods for data analysis. 3rd edition. Chapman & Hall/CRC; Boca Raton, FL: 2009.
- Chance FS, Abbott L, Reyes AD. Gain modulation from background synaptic input. *Neuron*. 2002; 35:773–782. [PubMed: 12194875]
- Chichilnisky EJ. A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*. 2001; 12:199–213. PMID: 11405422.

- Christianson GB, Sahani M, Linden JF. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *The Journal of Neuroscience*. 2008; 28:446–455. [PubMed: 18184787]
- Dahmen JC, Keating P, Nodal FR, Schulz A, King AJ. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*. 2010; 66:937–948. [PubMed: 20620878]
- David SV, Mesgarani N, Fritz JB, Shamma SA. Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *The Journal of Neuroscience*. 2009; 29:3374–3386. [PubMed: 19295144]
- Dean I, Harper NS, McAlpine D. Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*. 2005; 8:1684–1689.
- deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. *Science*. 1998; 280:1439–1444. [PubMed: 9603734]
- Deneve S, Latham PE, Pouget A. Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*. 1999; 2:740–745.
- Enroth-Cugell C, Shapley RM. Adaptation and dynamics of cat retinal ganglion cells. *The Journal of Physiology*. 1973; 233:271–309. [PubMed: 4747229]
- Escabi MA, Schreiner CE. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *The Journal of Neuroscience*. 2002; 22:4114–4131. [PubMed: 12019330]
- Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Efficiency and ambiguity in an adaptive neural code. *Nature*. 2001; 412:787–792. [PubMed: 11518957]
- Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*. 2003; 6:1216–1223.
- Geweke, J. *Bayesian Statistics*. Vol. Vol. 4. Oxford University Press; Oxford: 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments; p. 169-193.
- Gill P, Zhang J, Woolley S, Fremouw T, Theunissen FE. Sound representation methods for spectro-temporal receptive field estimation. *Journal of Computational Neuroscience*. 2006; 21:5–20. [PubMed: 16633939]
- Gourévitch B, Noreña A, Shaw G, Eggermont JJ. Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. *Cerebral Cortex*. 2009; 19:1448–1461. [PubMed: 18854580]
- Heeger DJ. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*. 1992; 9:181–197. PMID: 1504027. [PubMed: 1504027]
- Heeger DJ, Simoncelli EP, Movshon JA. Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences*. 1996; 93:623–627.
- Heuer HW, Britten KH. Contrast dependence of response normalization in area MT of the rhesus macaque. *Journal of Neurophysiology*. 2002; 88:3398–3408. [PubMed: 12466456]
- Jones, E.; Oliphant, T.; Peterson, P., et al. *SciPy: Open source scientific tools for Python*. 2001.
- Joris P, Yin T. Responses to amplitude-modulated tones in the auditory nerve of the cat. *The Journal of the Acoustical Society of America*. 1992; 91:215–232. PMID: 1737873. [PubMed: 1737873]
- Katzner S, Busse L, Carandini M. GABA-A inhibition controls response gain in visual cortex. *The Journal of Neuroscience*. 2011; 31:5931–5941. [PubMed: 21508218]
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of Computational Neuroscience*. 2000; 9:85–111. PMID: 10946994. [PubMed: 10946994]
- Kvale MN, Schreiner CE. Short-term adaptation of auditory receptive fields to dynamic stimuli. *Journal of Neurophysiology*. 2004; 91:604–612. [PubMed: 14762146]
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *Journal of Neurophysiology*. 2003; 90:2660–2675. [PubMed: 12815016]
- Liu, Bh; Wu, GK.; Arbuckle, R.; Tao, HW.; Zhang, LI. Defining cortical frequency tuning with recurrent excitatory circuitry. *Nature Neuroscience*. 2007; 10:1594–1600.

- Luo SX, Axel R, Abbott LF. Generating sparse and selective third-order responses in the olfactory system of the fly. *Proceedings of the National Academy of Sciences*. 2010; 107:10713–10718.
- Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. *The Journal of Neuroscience*. 2004; 24:1089–1100. PMID: 14762127. [PubMed: 14762127]
- Malmierca MS, Cristaudo S, Pérez-González D, Covey E. Stimulus-Specific adaptation in the inferior colliculus of the anesthetized rat. *The Journal of Neuroscience*. 2009; 29:5483–5493. [PubMed: 19403816]
- Malone BJ, Scott BH, Semple MN. Temporal codes for amplitude contrast in auditory cortex. *The Journal of Neuroscience*. 2010; 30:767–784. [PubMed: 20071542]
- Miller EK, Gochin PM, Gross CG. Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research*. 1993; 616:25–29. [PubMed: 8358617]
- Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology*. 2002; 87:516–527. [PubMed: 11784767]
- Missal M, Vogels R, Orban GA. Responses of macaque inferior temporal neurons to overlapping shapes. *Cerebral Cortex*. 1997; 7:758–767. [PubMed: 9408040]
- Moeller CK, Kurt S, Happel MFK, Schulze H. Longrange effects of GABAergic inhibition in gerbil primary auditory cortex. *European Journal of Neuroscience*. 2010; 31:49–59. [PubMed: 20092555]
- Murphy BK, Miller KD. Multiplicative gain changes are induced by excitation or inhibition alone. *The Journal of Neuroscience*. 2003; 23:10040–10051. [PubMed: 14602818]
- Nagel KI, Doupe AJ. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron*. 2006; 51:845–859. [PubMed: 16982428]
- Nelken I, Bizley J, Nodal F, Ahmed B, Schnupp J, King A. Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. *Journal of neurophysiology*. 2004; 92:25742588.
- Nelson PC, Carney LH. Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *Journal of Neurophysiology*. 2007; 97:522–539. [PubMed: 17079342]
- Ohshiro T, Angelaki DE, DeAngelis GC. A normalization model of multisensory integration. *Nature Neuroscience*. 2011; 14:775–782.
- Olsen SR, Bhandawat V, Wilson RI. Divisive normalization in olfactory population codes. *Neuron*. 2010; 66:287–299. [PubMed: 20435004]
- Olsen SR, Bortone DS, Adesnik H, Scanziani M. Gain control by layer six in cortical circuits of vision. *Nature*. 2012
- Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996; 381:607–609. [PubMed: 8637596]
- Patil A, Huard D, Fonnesbeck C. PyMC: Bayesian stochastic modelling in Python. *Journal of Statistical Software*. 2010; 35:1–81. [PubMed: 21603108]
- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ. Contrast gain control in auditory cortex. *Neuron*. 2011; 70:1178–1191. [PubMed: 21689603]
- Reale RA, Brugge JF. Directional sensitivity of neurons in the primary auditory (AI) cortex of the cat to successive sounds ordered in time and space. *Journal of Neurophysiology*. 2000; 84:435–450. [PubMed: 10899217]
- Recanzone GH, Wurtz RH, Schwarz U. Responses of MT and MST neurons to one and two moving objects in the receptive field. *Journal of Neurophysiology*. 1997; 78:2904–2915. [PubMed: 9405511]
- Rees A, Møller AR. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hearing Research*. 1983; 10:301–330. [PubMed: 6874603]
- Ringach DL. Population coding under normalization. *Vision Research*. 2010; 50:2223–2232. [PubMed: 20034510]

- Rotman Y, Bar-Yosef O, Nelken I. Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. *Hearing Research*. 2001; 152:110–127. [PubMed: 11223286]
- Ruderman DL, Bialek W. Statistics of natural images: scaling in the woods. *Physical Review Letters*. 1994; 73:814–817. [PubMed: 10057546]
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque V1 receptive fields. *Neuron*. 2005; 46:945–956. [PubMed: 15953422]
- Sahani, M.; Linden, JF. *Advances in Neural Information Processing Systems*. Vol. 15. MIT Press; Cambridge MA: 2003. How linear are auditory cortical responses?; p. 125-132.
- Schneider DM, Woolley SMN. Extra-classical tuning predicts stimulus-dependent receptive fields in auditory neurons. *The Journal of Neuroscience*. 2011; 31:11867–11878. [PubMed: 21849547]
- Scholl B, Gao X, Wehr M. Level dependence of contextual modulation in auditory cortex. *Journal of Neurophysiology*. 2008; 99:1616–1627. [PubMed: 18216226]
- Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. *Nature Neuroscience*. 2001; 4:819–825.
- Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA. Temporal symmetry in primary auditory cortex: implications for cortical connectivity. *Neural Computation*. 2007; 19:583–638. PMID: 17298227. [PubMed: 17298227]
- Simoncelli, EP.; Paninski, L.; Pillow, J.; Schwartz, O. Characterization of neural responses with stochastic stimuli. In: Gazzaniga, M., editor. *The Cognitive Neurosciences III*. MIT Press; Cambridge, MA: 2004. p. 327-338.
- Simoncelli EP, Heeger DJ. A model of neuronal responses in visual area MT. *Vision Research*. 1998; 38:743–761. [PubMed: 9604103]
- Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of Neuroscience*. 2000; 20:2315–2331. [PubMed: 10704507]
- Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*. 2004; 24:10440–10453. [PubMed: 15548659]
- Valentine PA, Eggermont JJ. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. *Hearing Research*. 2004; 196:119–133. [PubMed: 15464309]
- Wallace MN, Kitzes LM, Jones EG. Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex. *Experimental Brain Research*. 1991; 86:527–544. PMID: 1761089.
- Wark B, Fairhall A, Rieke F. Timescales of inference in visual adaptation. *Neuron*. 2009; 61:750–761. [PubMed: 19285471]
- Wehr M, Zador AM. Synaptic mechanisms of forward suppression in rat auditory cortex. *Neuron*. 2005; 47:437–445. [PubMed: 16055066]
- Winer JA, Miller LM, Lee CC, Schreiner CE. Auditory thalamocortical transformation: structure and function. *Trends in Neurosciences*. 2005; 28:255–263. [PubMed: 15866200]
- Woolley SMN, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*. 2005; 8:1371–1379.
- Woolley SMN, Gill PR, Theunissen FE. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *The Journal of Neuroscience*. 2006; 26:2499–2512. [PubMed: 16510728]
- Wu GK, Arbuckle R, Liu Bh, Tao HW, Zhang LI. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron*. 2008; 58:132–143. [PubMed: 18400169]
- Wu MC, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. *Annual Review of Neuroscience*. 2006; 29:477–505.
- Zhang J, Nakamoto KT, Kitzes LM. Modulation of level response areas and stimulus selectivity of neurons in cat primary auditory cortex. *Journal of Neurophysiology*. 2005; 94:2263–2274. [PubMed: 15917317]

- Zhu C, Byrd R, Lu P, Nocedal J. Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. *ACM: Transactions on Mathematical Software*. 1997; 23:550–560.
- Zoccolan D, Cox DD, DiCarlo JJ. Multiple object response normalization in monkey inferotemporal cortex. *The Journal of Neuroscience*. 2005; 25:8150–8164. [PubMed: 16148223]

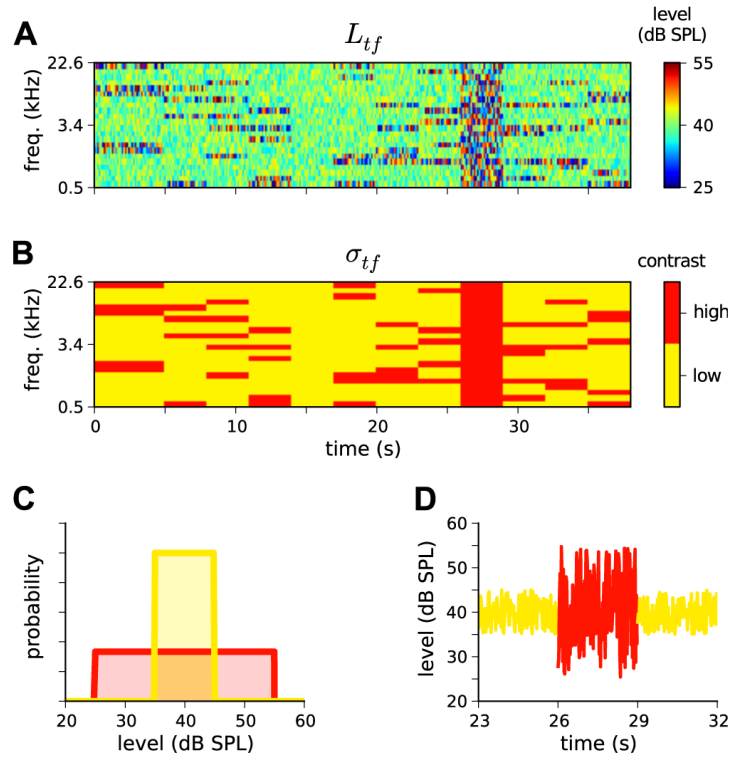


Figure 1.

Stimuli used to estimate contrast kernels, and their statistics. **A**, Schematic of a Random-Contrast Dynamic Random Chord (RC-DRC) stimulus. The stimulus comprises a sequence of chords, which change every 25 ms. The elements of the chords are pure tones, whose levels are drawn from one of the distributions shown in **C**. The color grid shows the sound level (L_{tf}) of a particular tone frequency at a particular time. **B**, The 38 s DRC stimulus shown in **A** is comprised of 12 segments in which the contrast in different frequency bins, σ_{tf} is either high (red) or low (yellow). **C**, Tone level distributions for low (yellow) and high (red) contrast segments. **D**, Level as a function of time for the 2.4 kHz tone over a 9 s period, i.e. a cross-section of **A**. This shows the transition from a segment where this tone's level distribution was low contrast (yellow), to a segment where it was high contrast (red), to a third segment where it was low contrast again (yellow).

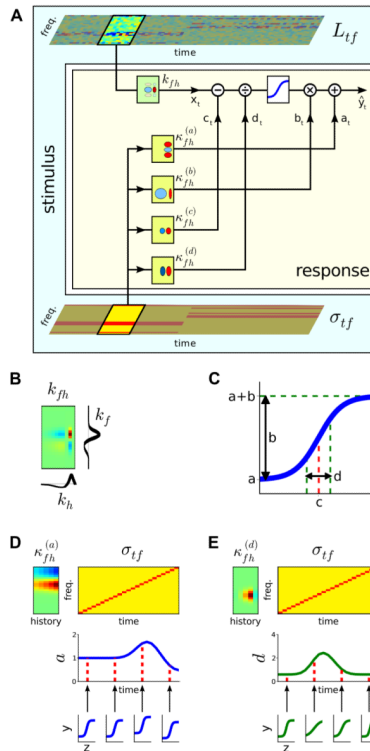


Figure 2.

Schematic of the contrast kernel model. **A**, The relationship between stimulus and neuronal response. The sound input is represented by its spectrogram, L_{tf} (top), and by its contrast profile, σ_{tf} (bottom). As in a standard linear-nonlinear model, the neural response is determined by convolving the spectrogram with a linear spectrotemporal kernel (k_{fh}), and passing the output of this operation (x_t) through a static output nonlinearity (here, a 4-parameter sigmoid, denoted by the blue curve), to produce the predicted spike rate (\hat{y}_t). The model developed here extends this by allowing each of the four parameters of the output nonlinearity (a through d , as shown in **C**) to change over time, depending on the statistics of recent stimulation. The evolution of each parameter $\theta \in \{a, b, c, d\}$ over time is determined

by convolving the contrast profile of the sound, σ_{tf} with a linear *contrast kernel*, $k_{fh}^{(\theta)}$. The effects of this on the shape of the output nonlinearity are illustrated in **D-E**. **B**, All spectrotemporal receptive fields and contrast kernels are assumed to be separable in

frequency and time, such that $k_{fh} = k_f \otimes k_h$, and $k_{fh}^{(\theta)} = k_f^{(\theta)} \otimes k_h^{(\theta)}$. This allows contrast kernels to be fitted in two stages: the spectral component (SCKs) in Figures 3–6, and the temporal component (TCKs) in Figure 7. **C**, The parameters of a sigmoidal static nonlinearity: a , the minimum firing rate; b , the output dynamic range; c , the stimulus inflection point; d , the (inverse) gain. **D**, An illustration of the effect of a contrast kernel for the nonlinearity parameter a , which sets the minimum firing rate of the output nonlinearity. Top left: a contrast kernel $k_{fh}^{(a)}$ is shown. Top right: the contrast profile of an example stimulus. Middle right: as a result of changing contrast, the parameter a changes with time. Bottom right: the effective shape of the output nonlinearities at different times, due to the changing value of a . These shifts would be combined with the contrast-dependent changes to the other nonlinearity parameters, b , c , and d , such as shown in **E**. **E**, Effect of a contrast kernel for the nonlinearity parameter d , which sets the (inverse) gain of the output nonlinearity. This

neuron decreases its gain when there is high contrast anywhere within a relatively broad region demarcated by $K_{fh}^{(d)}$.

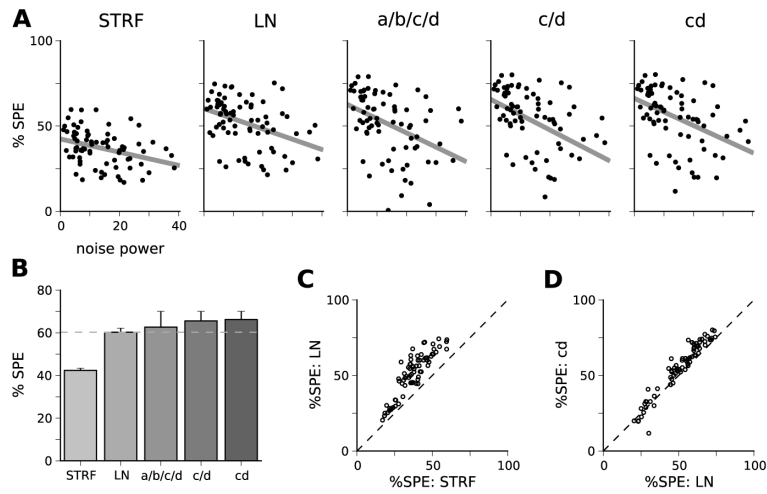


Figure 3.

Including spectral contrast kernels (SCKs) in models of neural responses improves their predictive power over the linear-nonlinear model; this is further improved by simplifying the model. **A**, Model predictive power, as measured in Sahani & Linden (2003). Model names are defined in Materials and Methods in the main text. For each model, scatter plots show the cross-validated prediction scores across all 77 units. These are calculated as the percentage of the unit's signal power (%SPE) captured by the model on the prediction dataset, and shown as a function of the normalized noise power in the unit's responses. Grey line shows the extrapolation of prediction scores to an idealized zero-noise unit, producing a lower bound on the model's overall predictive power over the population of auditory cortical units. The upper bound on predictive power has been omitted for clarity. **B**, Summary of predictive powers for the models in **A**. Solid bars show the lower bound (as plotted in **A**) from cross-validation; error bars show the upper bound from the training data set. Although adding a full set of contrast kernels (*a/b/c/d*) leads to a modest improvement in prediction scores over the linear-nonlinear (LN) model, the large number of parameters in the full model leads to overfitting. Rendering *a* and *b* contrast-independent reduces overfitting, and improves prediction scores (the *c/d* model). The best-performing model is the *cd* model, with a shared contrast kernel between *c* and *d*. **C**, Comparison between prediction scores for the LN model and for the STRF model, on a unit-by-unit basis. **D**, Comparison between the LN model and *cd*-model on a unit-by-unit basis.

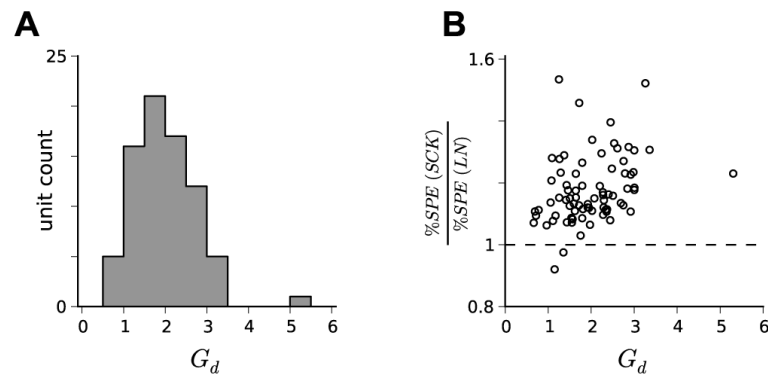


Figure 4.

Gain model: contrast-dependent gain changes across the population of A1/AAF units. **A**, The majority of units decreased their gain as contrast was increased, as expected. This is measured here by the ratio $G_d = d_{high}/d_{low}$. **B**, The larger a unit's contrast-dependent gain changes, the greater the improvement in model predictive power over the standard linear-nonlinear model. The (non-parametric) Spearman correlation coefficient between G_d and model improvement was 0.40 ($p < 0.001$).

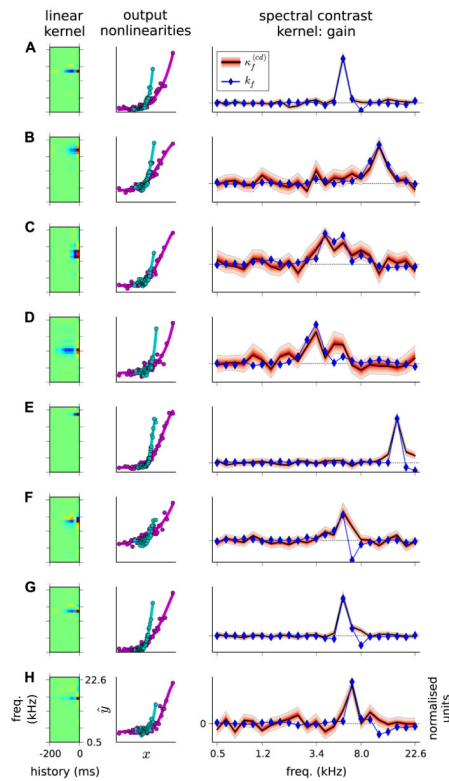


Figure 5.

Gain spectral contrast kernels (SCKs), for 8 example units. These are fits of the cd -model, with contrast-independent a and b , and a shared, real-valued SCK, $K_f^{(cd)}$, for c and d . Left: STRF for each unit. Middle: static output nonlinearities for each unit, when estimated under the all-high-contrast condition (magenta), and the all-low-contrast condition (cyan), showing the gain change between the two conditions. Right: SCK for each unit. The black line shows the MAP estimate for $K_f^{(cd)}$; the red filled region, bounded by the grey lines, shows a 95% credible interval for the posterior distribution over these coefficients. The red shading increases in darkness with probability. The blue line and blue diamonds show the frequency component of the linear, separable STRF, k_f . Both k_f and $K_f^{(cd)}$ have been normalized by the respective standard deviations to facilitate visual comparison. **A-D** exemplify how k_f and $K_f^{(cd)}$ align in best frequency and bandwidth. **E-G** (but not **H**) show examples where $K_f^{(cd)}$ covers the inhibitory sidebands of the receptive field.

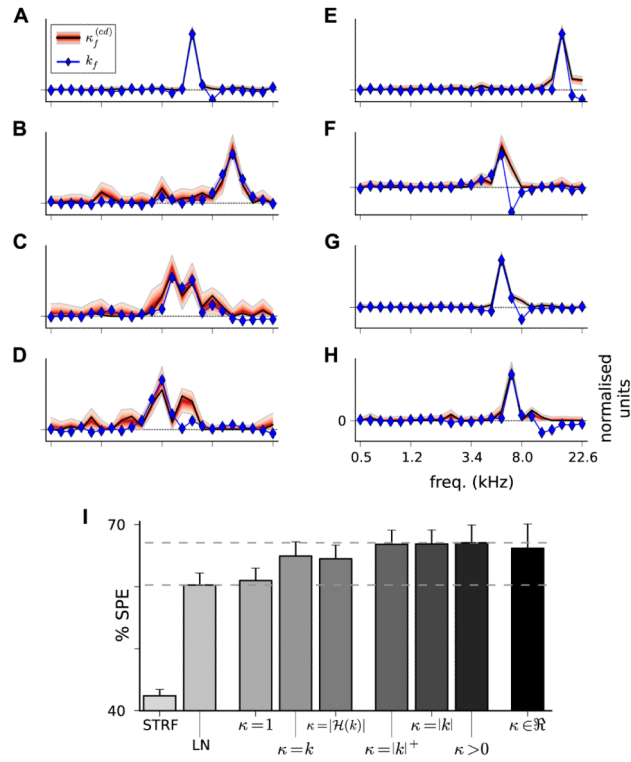


Figure 6. Approximations to the cd -model. **A-H**, Gain SCKs when coefficients were constrained to be positive. This shows the same 8 units as shown in Figure 5. Again, the frequency component of the STRF, k_f (blue), approximately matches the gain SCK, $\kappa_f^{(cd)}$ (black line and red area). **I**, Model predictive power for the cd -model with constrained coefficients; as in Figure 3B, solid bars show prediction scores, and error bars show training scores. When the contrast kernel coefficients are unconstrained ($\kappa \in \mathfrak{R}$; right), the model performance is better than the linear (STRF) and linear-nonlinear (LN) models (left). Restricting the coefficients of the SCK to be positive ($\kappa > 0$) reduces overfitting and improves prediction scores. Excellent approximations are provided by fixing the SCKs as either the absolute value of the STRF frequency kernel ($\kappa = |k|$), or the rectified value ($\kappa = |k|^+$). Models which do not perform as well include fixing the the contrast kernel as the STRF frequency kernel ($\kappa = k$), fixing it as the magnitude of the Hilbert transform of the STRF frequency kernel ($\kappa = |H(k)|$), or assuming that it is constant with respect to frequency ($\kappa = 1$). These still out-perform the simple linear-nonlinear (LN) model. Dashed lines are shown at the model performance values for the LN model and the constrained-positive cd -model.

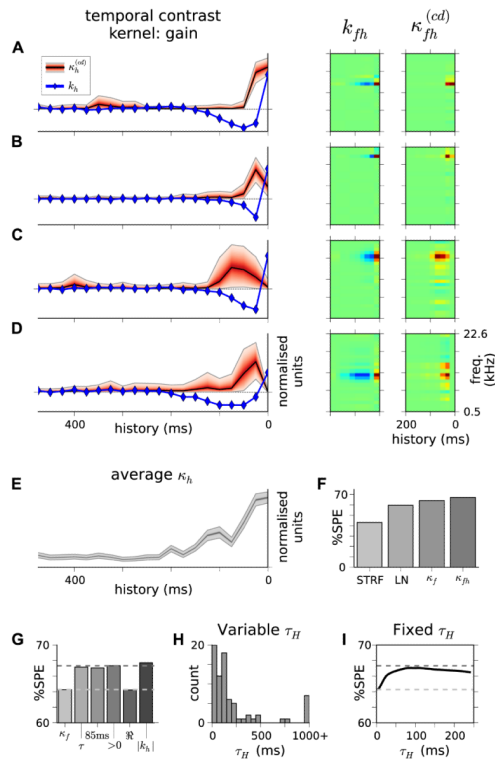


Figure 7.

Temporal contrast kernels (TCKs). **A-D**, Left panels show the TCKs for four example units.

As in Figures 5 and 6, red area shows the gain TCK, $\kappa_h^{(cd)}$, while blue line and diamonds show the temporal component of the STRF, k_{fh} . Right panels compare the STRF, k_{fh} , with

the full spectrotemporal contrast kernels, $\kappa_{fh}^{(cd)}$, as per Figure 2. **E**, Mean of the contrast time kernels from the 77 cortical units, $\kappa_h^{(cd)}$. This shows the approximately exponential shape of the time kernels. The mean contrast kernel had a fitted time constant of 86 ms. **F**, Model

predictive power. Including a history component to the contrast kernels (κ_{fh}) improves the performance of the model, compared with the assumption that only the current contrast matters (κ_f). Prediction scores for the simple STRF model and the linear-nonlinear (LN) model are shown for comparison. Note that this is fitted over a different dataset from that used in Figures 3–6, so the values of %SPE in this figure do not match those presented previously. **G**, Model predictive powers for a range of TCK models. In order, from left to right, these models are the following: (κ_f), no history dependence, i.e. $\kappa_h = \delta_{t0}$; (τ), exponential model with time constant τ_H fitted (see **H**); (85 ms), exponential model with τ_H fixed at 85 ms (see **I**); (> 0), κ_h constrained to be positive; (\Re), κ_h allowed to take on any real value; ($|k_{fh}|$), κ_h approximated as the absolute value of the STRF time kernel. Dashed horizontal lines show the model predictive power for the κ_f and the > 0 models. Note that allowing the coefficients of the TCK to be real-valued (the \Re model) led to considerable overfitting; the > 0 model is thus the STCK model considered in the main text (see Materials and Methods). **H**: Fits of the time constant τ_H for the exponential model for all 77 units. The median time constant was 117 ms. **I**: Model predictive power for the exponential model when τ_H was fixed, rather than fitted. Abscissa denotes the fixed value of τ_H , ordinate as in **G**. The horizontal dashed lines are as in **G**. The most predictive model had $\tau_H = 85$ ms. Thus, three different measures of the time course of gain changes (in **E**, **H**, and **I**) give roughly consistent answers.

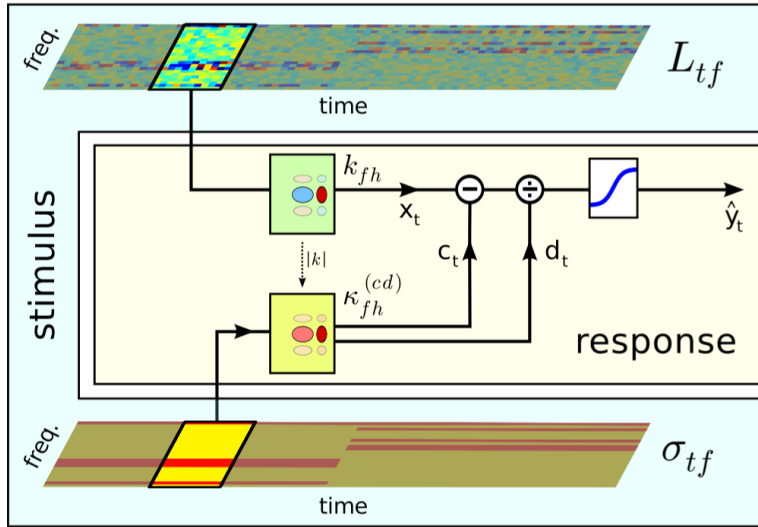


Figure 8. Summary of results. We find that the gain changes undergone by cortical neurons in response to complex patterns of stimulus contrast can be captured by this simplified contrast kernel model. The neural response is determined by convolving the spectrogram with a linear spectrotemporal kernel (k_{fh}), and passing the output of this operation (x_t) through a static output nonlinearity, to produce the predicted spike rate (\hat{y}_t). The minimum and maximum firing rate of the output nonlinearity are fixed, but the stimulus inflection point (c) and the (inverse) gain (d) change over time, depending on the statistics of recent stimulation. The evolution of c and d over time is determined by convolving the contrast profile of the sound, σ_{tf} , with a single contrast kernel, $\kappa_{fh}^{(cd)}$, as in Equation (11). Finally, the contrast kernel can be approximated as $\kappa_{fh}^{(cd)} \approx |k_{fh}|$. This model captures 20-25% of the residual variance not explained by the LN model by adding only a further two parameters.