

Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing

Ben D.B. Willmore,^{1*} Oliver Schoppe,^{1,2*} Andrew J. King,¹ Jan W.H. Schnupp,^{1**} and Nicol S. Harper^{1**}

¹Department of Physiology, Anatomy, and Genetics, University of Oxford, Oxford OX1 3PT, United Kingdom, and ²Bio-Inspired Information Processing, Technische Universität München, 85748 Garching, Germany

Adaptation to stimulus statistics, such as the mean level and contrast of recently heard sounds, has been demonstrated at various levels of the auditory pathway. It allows the nervous system to operate over the wide range of intensities and contrasts found in the natural world. Yet current standard models of the response properties of auditory neurons do not incorporate such adaptation. Here we present a model of neural responses in the ferret auditory cortex (the IC Adaptation model), which takes into account adaptation to mean sound level at a lower level of processing: the inferior colliculus (IC). The model performs high-pass filtering with frequency-dependent time constants on the sound spectrogram, followed by half-wave rectification, and passes the output to a standard linear–nonlinear (LN) model. We find that the IC Adaptation model consistently predicts cortical responses better than the standard LN model for a range of synthetic and natural stimuli. The IC Adaptation model introduces no extra free parameters, so it improves predictions without sacrificing parsimony. Furthermore, the time constants of adaptation in the IC appear to be matched to the statistics of natural sounds, suggesting that neurons in the auditory midbrain predict the mean level of future sounds and adapt their responses appropriately.

Key words: adaptation; auditory cortex; inferior colliculus; mean sound level; model; spectrotemporal receptive field

Significance Statement

An ability to accurately predict how sensory neurons respond to novel stimuli is critical if we are to fully characterize their response properties. Attempts to model these responses have had a distinguished history, but it has proven difficult to improve their predictive power significantly beyond that of simple, mostly linear receptive field models. Here we show that auditory cortex receptive field models benefit from a nonlinear preprocessing stage that replicates known adaptation properties of the auditory midbrain. This improves their predictive power across a wide range of stimuli but keeps model complexity low as it introduces no new free parameters. Incorporating the adaptive coding properties of neurons will likely improve receptive field models in other sensory modalities too.

Introduction

Adaptation to stimulus statistics is an important process in sensory coding, whereby neurons adjust their sensitivity in response

to the statistics of recently presented stimuli (Fairhall et al., 2001; Wark et al., 2007; Carandini and Heeger, 2012). For example, neurons in the auditory nerve (Wen et al., 2009), inferior colliculus (IC; Dean et al., 2008), and cortex (Watkins and Barbour, 2008; Rabinowitz et al., 2013) shift their dynamic ranges to compensate for changes in the mean level of recent sound stimulation. Neurons in the auditory periphery (Joris and Yin, 1992), midbrain (Rees and Møller, 1983; Kvale and Schreiner, 2004; Dean et al., 2005; Nelson and Carney, 2007; Dahmen et al., 2010; Rabinowitz et al., 2013), and higher auditory pathways (Nagel

Received June 26, 2015; revised Nov. 3, 2015; accepted Nov. 10, 2015.

Author contributions: B.D.B.W. and N.S.H. designed research; B.D.B.W., O.S., and N.S.H. performed research; A.J.K. contributed unpublished reagents/analytic tools; B.D.B.W., O.S., and N.S.H. analyzed data; B.D.B.W., O.S., A.J.K., J.W.H.S., and N.S.H. wrote the paper.

This work was supported by the Wellcome Trust through a Principal Research Fellowship to A.J.K. (WT076508AIA). O.S. was supported by the German National Academic Foundation. N.S.H. was supported by a Sir Henry Wellcome Postdoctoral Fellowship (WT082692) and other Wellcome Trust funding (WT076508AIA); by the Department of Physiology, Anatomy, and Genetics at the University of Oxford; by Action on Hearing Loss Grant PA07; and by Biotechnology and Biological Sciences Research Council Grant BB/H008608/1. We are grateful to Josh McDermott, Neil Rabinowitz, Sandra Tolnai, Kerry Walker, Fernando Nodal, Astrid Klinge-Strahl, Vani Rajendran, and James Cooke for help with these experiments and to Stephen David for providing stimulus generation code. We used sounds from the freesound.org database contributed by users juskiddink, felix-blume, duckduckpony, lebractflp, and spoonbender.

The authors declare no competing financial interests.

*B.D.B.W. and O.S. are joint first authors.

**J.W.H.S. and N.S.H. joint last authors.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to Ben D.B. Willmore, Department of Physiology, Anatomy, and Genetics, University of Oxford, Sherrington Building, Parks Road, Oxford OX1 3PT, UK. E-mail: benjamin.willmore@dpag.ox.ac.uk

DOI:10.1523/JNEUROSCI.2441-15.2016

Copyright © 2016 Willmore, Schoppe et al.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

and Doupe, 2006; Malone et al., 2010; Rabinowitz et al., 2011) also adapt to the variance of recently presented stimuli. Similar processes operate in the visual (Mante et al., 2005) and somatosensory (Garcia-Lazaro et al., 2007) systems, and it has been proposed that these forms of adaptation allow the nervous system to efficiently represent stimuli across the wide range of intensities and contrasts found in the natural world (Fairhall et al., 2001).

Functional models aimed at predicting responses of sensory neurons generally do not incorporate adaptation to stimulus statistics. In the auditory system, such models typically involve variations of the spectrotemporal receptive field (STRF), the standard computational model of neuronal responses (Aertsen et al., 1980; Aertsen and Johannesma, 1981; deCharms et al., 1998; Klein et al., 2000; Theunissen et al., 2000; Escabí and Schreiner, 2002; Miller et al., 2002; Fritz et al., 2003; Linden et al., 2003; Gill et al., 2006; Christianson et al., 2008; David et al., 2009; Gourevitch et al., 2009). Each STRF is a set of coefficients that describe the best linear approximation to the relationship between the spiking responses of a neuron and the power in the spectrogram of the sounds heard by the animal.

In principle, STRFs are powerful computational tools because they provide both a way to characterize neurons, by quantifying their sensitivity to different sound frequencies, and to predict responses to arbitrary new stimuli (deCharms et al., 1998; Schnupp et al., 2001; Escabí and Schreiner, 2002). In practice, STRFs are only moderately successful in achieving this (Linden et al., 2003; Machens et al., 2004). To improve the predictive power of STRFs, nonlinear extensions have been proposed, including output nonlinearities (Atencio et al., 2008; Rabinowitz et al., 2011), feedback kernels (Calabrese et al., 2011), second-order interactions, and input nonlinearities (Ahrens et al., 2008; David et al., 2009; David and Shamma, 2013). However, prediction accuracy remains far from perfect. Also, some of these approaches add complexity to the model and can be difficult to interpret in biological terms. Here we take an alternative approach that seeks to improve the prediction accuracy of STRF-like models by incorporating a simple, adaptive, nonlinear preprocessing step that mimics the physiological properties of neurons in the auditory midbrain.

Adaptation to mean sound level in the IC has been characterized by Dean et al. (2008), who measured how the time constants of adaptation in guinea pigs vary with frequency. This information can be used to build a model of adaptation to stimulus statistics in the IC, which can then be incorporated into an STRF model of neural responses. We recorded the responses of neurons in ferret auditory cortex to a range of sounds and constructed STRF models relating the responses to the sound spectrograms. We then augmented these models by incorporating a nonlinear transform of the spectrogram, which captures adaptation to mean sound level in the IC. Since the IC provides an obligatory relay for ascending inputs to the auditory cortex, this transform was incorporated at the input stage of the model, forming a nonlinear-linear-nonlinear (NLN) cascade. This NLN model provides a substantial improvement over the standard STRF models in describing and predicting the responses of cortical neurons.

Materials and Methods

Experimental procedures

All animal procedures were approved by the local ethical review committee and performed under license from the UK Home Office. Ten adult pigmented ferrets (seven female, three male; all >6 months of age) underwent electrophysiological recordings under ketamine–medetomidine anesthesia. Full details are as in the study by Bizley et al. (2009). Briefly,

we induced general anesthesia with a single intramuscular dose of medetomidine ($0.022 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$) and ketamine ($5 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), which was then maintained with a continuous intravenous infusion of medetomidine and ketamine in saline. Oxygen was supplemented with a ventilator, and we monitored vital signs (body temperature, end-tidal CO_2 , and the electrocardiogram) throughout the experiment. The temporal muscles were retracted, a head holder was secured to the skull surface, and a craniotomy and a durotomy were made over the auditory cortex. We made extracellular recordings from neurons in primary auditory cortex (A1) and the anterior auditory field (AAF) using silicon probe electrodes (Neuronexus Technologies) with 16 or 32 sites (spaced at 50 or 150 μm) on probes with one, two, or four shanks (spaced at 200 μm). Stimuli were presented via Panasonic RPHV27 earphones, which were coupled to otoscope specula that were inserted into each ear canal, and driven by Tucker-Davis Technologies System III hardware (48 kHz sample rate). We clustered spikes off-line using *klustakwik* (Kadir et al., 2014); for subsequent manual sorting, we used either *spikemonger* (an in-house package) or *klustaviewa* (Kadir et al., 2014).

Stimuli

We used several stimulus classes: two types of dynamic random chords (DRCs), temporally orthogonal ripple combinations (TORCs), modulated noise, and natural sounds.

DRCs (deCharms et al., 1998; Schnupp et al., 2001; Rutkowski et al., 2002; Linden et al., 2003) consist of sequences of superposed pure tones whose levels are chosen pseudorandomly. Each chord contained 31 pure tones whose frequencies were log-spaced between 1 kHz and 32 kHz at 1/6 octave intervals. Each chord lasted 62.5 ms with 5 ms linear ramps between chords. The levels were chosen from a uniform distribution between 30 and 70 dB sound pressure level (SPL). We also included variable-rate DRCs, a novel stimulus designed to have a richer modulation structure, while retaining the other advantages of DRCs. In this case, each chord lasted 10.4 ms, but the level of each tone was kept constant for between 1 and 12 chords (lengths were chosen from a uniform distribution, independently for each frequency), rather than changing on every chord.

TORCs (Klein et al., 2000) consist of superposed noise stimuli with spectrograms modulated by superpositions of sinusoids. We used a set of 30 TORCs (each 3 s long) covering frequency space from 1 to 32 kHz, with temporal modulations from 4 to 48 Hz and frequency modulations up to 1.4 cycles/octave.

The modulated noise stimulus was generated using the sound texture synthesis algorithm developed by McDermott and Simoncelli (2011). The modulated noise had a pink power spectrum between 1 and 32 kHz and a white modulation spectrum between 13.3 and 160 Hz. This stimulus has a somewhat naturalistic structure, but without the complex higher-order statistical relationships of real, natural sounds.

In the first series of experiments (BigNat), we presented natural sounds only. We made recordings from 535 units in six ferrets (five female, one male). There were 20 sound clips of 5 s duration each, separated from each other by ~ 0.25 s silence. We recorded responses to the clips, presented in random order and repeated this 20 times. The sound clips included recordings of animal vocalizations (e.g., ferrets and birds), environmental sounds (e.g., water and wind), and speech. The sequences had root mean square intensities in the range 75–82 dB SPL. We presented the sounds at a sampling rate of 48,828.125 Hz. We discarded data recorded in the first 250 ms after the onset of each stimulus, leaving an effective data size of 20×95 s (20 repeats of 20 sounds with a duration of 4.75 s each).

In the second series of experiments (Comparison), we presented natural sounds, DRCs, TORCs, and modulated noise in an interleaved fashion, to enable comparison between different stimulus types. We recorded responses to the clips, presented in random order, and repeated this 10 times. Recordings were made from 220 units in four ferrets (two female, two male). The stimulus sampling rate was 97,656.25 Hz. Again, we discarded the first 250 ms after the onset of each stimulus, leaving an effective data size of $5 \times 10 \times 45$ s (five stimulus types with 10 repeats of 45 s each).

In the Comparison dataset, the natural sounds were 1 s snippets of vocalizations (human, bird, sheep) and environmental sounds. These

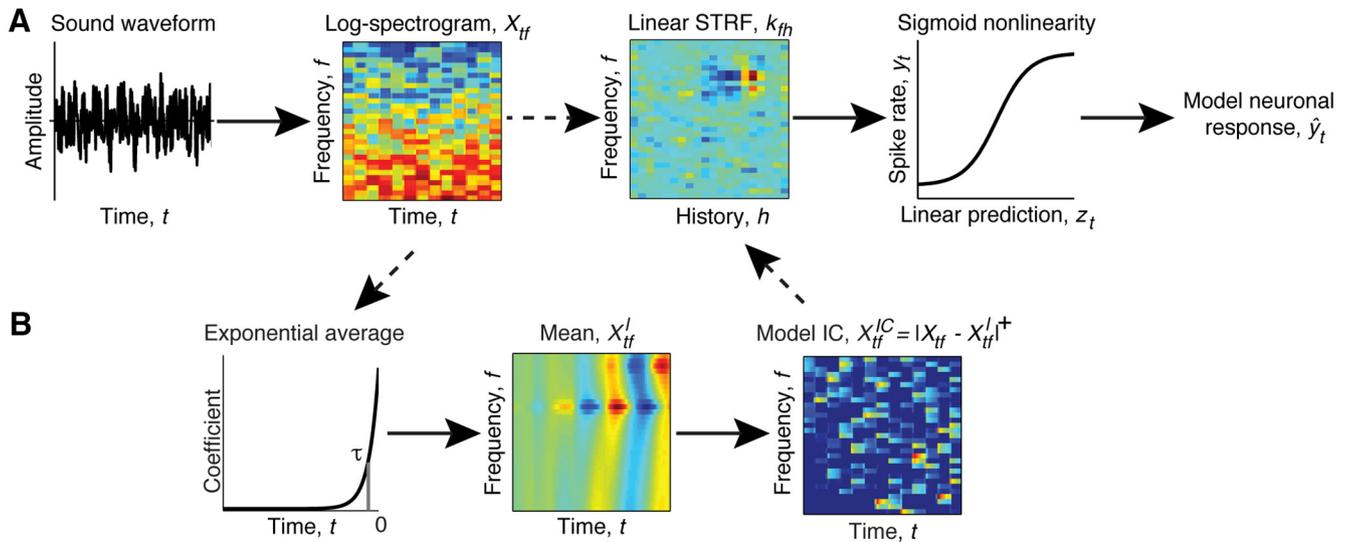


Figure 1. Two models of the stimulus–response relationship for auditory neurons. **A**, Standard LN model. The log-spectrogram of the sound waveform, X_{tf} , is operated on by a linear kernel, k_{fh} , and sigmoid output nonlinearity to produce a model, \hat{y}_t , of the neuronal response. **B**, The IC Adaptation model augments the LN model by adding a nonlinear transform of the spectrogram. The dashed arrows indicate the alternative processing paths for the Standard LN and IC Adaptation models. The nonlinear transform consists of high-pass filtering each frequency band of the spectrogram by subtracting the convolution of that frequency band with an exponential filter with time constant τ (shown by the vertical line), followed by half-wave rectification. The resulting modified spectrogram, X_{tf}^C , is then used as an alternative input to the standard LN model.

were separated by silent gaps, and the silent periods along with the first 250 ms of neural responses after each silent period were removed.

Neural responses

For each unit, we counted spikes in 5 ms time bins and averaged these counts over all trials to compute the peristimulus time histogram (PSTH). We smoothed the PSTH with a 21 ms Hanning window (Hsu et al., 2004) to estimate each neuron’s evoked firing rate. We denote the (trial-averaged) neuronal response as y_t .

Unit selection criterion

Only units whose firing rate was modulated in response to the stimuli in a reliable, repeatable manner were included for analysis. We measured this using the noise ratio (NR; Sahani and Linden, 2003; Rabinowitz et al., 2011) for the PSTH of each unit:

$$\text{noise ratio} = \frac{\text{noise power}}{\text{signal power}} = \frac{\text{total variance} - \text{explainable variance}}{\text{explainable variance}}.$$

Each unit was included in our analyses if it had a noise ratio of <200 across the entire dataset (i.e., across all natural stimuli for the BigNat set or across all stimulus classes for the Comparison set). Three hundred of 535 units were included from the BigNat set and 77 of 220 from the Comparison set.

Log-spectrograms

We characterized the power in each stimulus using a log-spaced, log-valued spectrogram. We first calculated the spectrogram of each sound using 10 ms Hanning windows, overlapping by 5 ms (giving 5 ms temporal resolution). We then aggregated across frequency using overlapping triangular windows with log-spaced characteristic frequencies to compute the signal power in each frequency band, using code modified from melbank.m by Mike Brookes (Imperial College London, London, UK; <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>). For the BigNat stimulus, we used 34 log-spaced frequencies from 500 to 22.627 Hz (1/6 octave spacing). For the Comparison stimuli, we used 31 log-spaced frequencies from 1 to 32 kHz (also 1/6 octave spacing). Finally, we took the logarithm of the resulting values, and values lower than a threshold (approximately equivalent to the mean activity caused by a 0 dB SPL flat spectrum noise) were set to that threshold, giving the log-spectrogram, X_{tf} , at time t and frequency f .

As input to the models, we reorganized X_{tf} as a three-tensor X_{tfh} , where X_{tfh} gives the sound intensity (elements of X_{tf}) for the recent stimulus

history, $h = 0$ to $H-1$ time bins in the past, from time t , at frequency f , i.e., $X_{tfh} = X_{(t-h),f}$. For STRF estimation, we used a history length of 20 bins of 5 ms duration (100 ms total).

Model testing and comparison

To fit and test our models, we used a k -fold testing procedure ($k = 10$) for both datasets. Thus, each dataset was split into 10 segments consisting of a contiguous 10% of the data. One of the 10 segments was set aside as a test set, and the model was trained on the remaining 90% of the data (the training set) to fit the STRF and the parameters of the nonlinearity. Model performance was then measured with the unused test set, i.e., the model was used to predict the neural response to the test set stimulus. We repeated this process 10 times, each time using a different segment as a test set, and averaged the performance measure over the 10 segments.

Linear–nonlinear STRF model

We described the responses of cortical neurons using two models. The first was a standard linear–nonlinear (LN) model (Chichilnisky, 2001; Simoncelli et al., 2004; Fig. 1A) relating neural responses, y_t , to the log-spectrogram, X_{tf} , of the stimuli. To do this, we first found the STRF k_{fh} , the linear approximation to the mapping between the PSTH, y_t , and the log-spectrogram, X_{tf} . We estimated k_{fh} by minimizing (subject to regularization; see below) the mean squared error between the PSTH, y_t , and its linear estimate from X_{tf} . This linear estimate, z_t , is given by the following:

$$z_t = \sum_{f,h} X_{tfh} k_{fh}. \quad (1)$$

Previous studies have used separable kernel estimation (Linden et al., 2003; Rabinowitz et al., 2011), which sometimes provides better descriptions of auditory neurons than inseparable approaches (particularly using DRC stimuli). Here, however, we used inseparable kernels to allow for the possibility of inseparable kernel structure with the TORCs and modulated noise stimuli. To estimate inseparable kernels, we used glmnet for Matlab (J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon, Stanford University, Stanford, CA; see http://web.stanford.edu/~hastie/glmnet_matlab/), which uses elastic net regularization. This technique can optimize k_{fh} using a linear combination of L1 (Willmore et al., 2010) and L2 (Willmore and Smyth, 2003) penalties we have used in the past, with a parameter, α , that determines the relative strength of each penalty. We explored three approaches: using an L1 penalty ($\alpha = 1$), using an L2 penalty ($\alpha = 0$), and optimizing α for each unit. Here we

present the results obtained using L2 regularization, but the results are similar for the other forms of regularization. The regularization parameter, λ , determines the strength of regularization. To determine the optimal choice of λ , we reserved a randomly chosen 10% of each training set for cross-validation. STRFs were estimated using the remaining 90% of the training set, using a wide range of choices of λ . We then selected the STRF that provided the best prediction (minimum mean square error) on the cross-validation set. The use of three separate subsets (where STRFs are fitted using one set, the regularization parameter is chosen using a second set, and prediction scores are measured using a third set), minimizes overfitting in both model fitting and assessment.

We then fitted a sigmoid (logistic) nonlinearity to relate the output of the linear model, z_t , to the neural responses by minimizing the mean squared error between the PSTH, y_t , and the nonlinear estimate of the PSTH, \hat{y}_t :

$$\hat{y}_t = a + \frac{b}{1 + \exp(-(z_t - c)/d)}, \quad (2)$$

where a is the minimum firing rate, b is the output dynamic range, c is the input inflection point, and d is the reciprocal of the gain (Rabinowitz et al., 2011, 2012). All parameters a , b , c , and d were fitted to the whole training set, using minFunc by Mark Schmidt (University of British Columbia, British Columbia, Canada; <http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html>).

To ensure that this sequential fitting procedure did not adversely affect our results, we also fitted the STRF and the sigmoid output nonlinearity using two other fitting methods (for the Comparison dataset only). The first was an iterative procedure where we estimated the STRF, then estimated the sigmoid, then inverted the sigmoid and refitted the STRF, and repeated for 10 iterations. The second was a neural network with linear input units, one logistic hidden unit, and a final linear output unit; this model was fitted using backpropagation. Both of these models are identical in mathematical form to the original model and only differ in the fitting procedure. We found that they provided very similar results (data not shown) to the sequential fitting procedure.

Nonlinear–linear–nonlinear STRF models

To extend the LN model to incorporate our knowledge about adaptation to mean sound level in the IC, we introduced a nonlinear transformation of the log-spectrogram, producing the IC Adaptation model (Fig. 1B). To test the importance of different aspects of this model, we also used several variations as controls.

IC Adaptation model. We convolved every frequency band in the log-spectrogram of the stimulus with an exponential filter, E_{fh} :

$$X_{tf}^1 = \sum_h X_{(t-h),f} E_{fh}, \text{ where } E_{fh} = \frac{1}{N_f} \exp(-h/\tau_f). \quad (3)$$

N_f was a normalization constant, chosen so that the exponential filter for each frequency band summed to 1. Here, the number of time bins, H , is 499, giving 2.5 s of history. The time constants, τ_f , of the filters varied with sound frequency, following the frequency dependence of the time constant found by Dean et al. (2008). The relationship we used was a linear regression (Fig. 2) relating τ_f (in milliseconds) to the logarithm of the units' characteristic frequency (in hertz):

$$\tau_f = 500 - 105 \log_{10}(f), \quad (4)$$

so that τ_f depends on the logarithm of frequency, between $\tau_f = 500$ Hz = 217 ms and $\tau_f = 32$ kHz = 27 ms.

The time-varying response of each exponential filter, X_{tf}^1 , was then subtracted from the corresponding frequency band in the log-spectrogram, X_{tf} , giving a high-pass-filtered version, $X_{tf}^h = X_{tf} - X_{tf}^1$, which was then half-wave rectified to give $X_{tf}^{1C} = |X_{tf}^h|^+$. We then used X_{tf}^{1C} in place of X_{tf} for STRF analysis.

No-half-wave-rectification model. This model is the same as the IC Adaptation model, but without half-wave rectification; i.e., X_{tf}^h was used for STRF analysis.

Median- τ model (τ^{med}). This model is the same as the IC Adaptation model, but a fixed time constant [$\tau^{\text{med}} = 160$ ms; equal to the median of

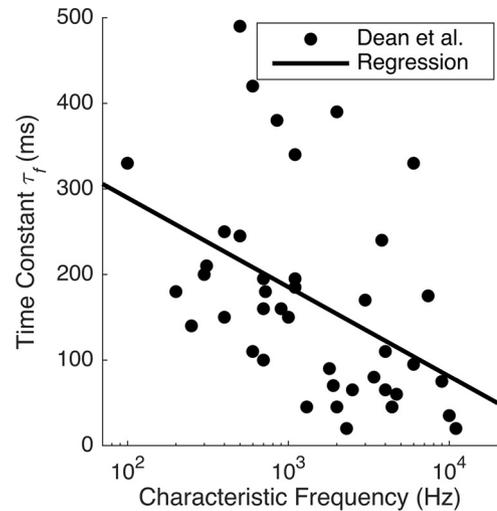


Figure 2. Time constants of adaptation to stimulus mean observed by Dean et al. (2008) in the guinea pig IC. The x -axis shows the characteristic frequency of each IC unit, and the y -axis shows the time constant of an exponential fit to the adaptation curve for the corresponding unit. The line is our regression fit to these data (Eq. 4).

the time constants measured by Dean et al. (2008)] was used for all frequency channels instead of the frequency-dependent τ_f .

Minimum- τ model (τ^{min}). This model is the same as the IC Adaptation model, but with $\tau^{\text{min}} = 27$ ms.

Maximum- τ model (τ^{max}). This model is the same as the IC Adaptation model, but with $\tau^{\text{max}} = 217$ ms.

Performance measures

Prediction performance was primarily assessed using the normalized correlation coefficient (CC_{norm}), as introduced for coherence by Hsu et al. (2004), and used for the correlation coefficient by Touryan et al. (2005). The prediction accuracy, as quantified by the raw correlation coefficient, CC_{raw} , is affected both by model performance and by the variability of neural responses to the stimulus. To correct for the contribution of response variability, and measure only model performance, we use CC_{norm} , defined as the ratio of the CC_{raw} to the theoretical maximum CC_{max} :

$$CC_{\text{norm}} = \frac{CC_{\text{raw}}}{CC_{\text{max}}}. \quad (5)$$

CC_{max} is the correlation coefficient between the recorded mean firing rate across all repeats of the stimulus and the (unknown) true mean firing rate (measured over infinite repeats) and is an upper limit on model performance. Following Hsu et al. (2004) and Touryan et al. (2005), we estimated CC_{max} using the following:

$$CC_{\text{max}} = \sqrt{\frac{2}{1 + 1/CC_{\text{half}}}}, \quad (6)$$

where CC_{half} is the correlation coefficient of the mean PSTH for one-half of the trials with the mean PSTH for the other half of the trials. We took the mean CC_{half} over all 126 possible combinations for the Comparison dataset (10 trials) and over 126 randomly chosen sets of half of the trials for the BigNat dataset (20 trials).

Natural sound analysis

In addition to modeling the neural responses, we performed an analysis of a database of natural sound recordings with the aim of asking how the distribution of IC adaptation time constants reported in the literature across frequencies (Dean et al., 2008) might relate to the properties of the natural acoustic environment. Most sounds in the database were recorded by us under various conditions (ranging from open air to an anechoic chamber). An additional seven sound recordings were taken from the freesound.org database. We arranged the sounds into seven

broad categories: breaking wood sounds, crackling fire, rustling foliage, vocalizations (human, frog, bird, sheep), walking footsteps on numerous surfaces, ocean and river water (“water”), and rain and thunderstorms (“weather”). We calculated the log-spectrogram, X_{tf} , of each sound (as for the STRF analysis) and estimated the time-varying mean level in each frequency band, μ_{tf} , by convolution with a boxcar filter of length, T_{av} (for eight log-spaced values of T_{av} between 15 and 1000 ms), arranged so that μ_{tf} at time t contained the mean sound level between t and $t + T_{av}$. Thus, μ_{tf} is an estimate of the mean sound level in a given frequency band in the immediate future.

We concatenated all sounds in a given category (including, at most, 10 s from any single sound). We then estimated a set of linear filters, E_{fh}^{nat} (one per frequency band, sound category, and value of T_{av}), which operated over 2.5 s of sound history and were optimized to produce an estimate, $\hat{\mu}_{tf}$, of the time-varying mean level, μ_{tf} :

$$\hat{\mu}_{tf} = \sum_h X_{(t-h),f} E_{fh}^{nat}. \quad (7)$$

The kernels, E_{fh}^{nat} , were constrained to be exponential in shape as follows:

$$E_{fh}^{nat} = A_f \exp(-h/\tau_f^{nat}). \quad (8)$$

Estimating the kernels therefore consisted of fitting two parameters: A_f (amplitude) and τ_f^{nat} (time constant). We optimized these parameters to minimize the mean squared error between $\hat{\mu}_{tf}$ and μ_{tf} using `fminsearch` in Matlab. Both parameters were allowed to vary freely for different frequencies.

Results

Using the methods described above, we presented a range of natural and synthetic sounds to anesthetized ferrets and recorded responses of neurons in the primary cortical areas A1 and AAF. We modeled these responses using a classical LN model of spectrotemporal tuning as well as using a novel model that included a nonlinear input stage that incorporates adaptation of IC neurons to stimulus statistics (the IC Adaptation model). We compared the predictive power of the two models by measuring the accuracy of their prediction of neural responses to a reserved test set of sounds.

The IC Adaptation model predicts responses to natural sounds more accurately than conventional LN models

We first evaluated the performance of the IC Adaptation model on the responses of each unit to a set of natural sounds (BigNat dataset). Natural sounds provide the ultimate test of models of neural responses because of their ecological relevance. A good model should be able to predict responses to natural sounds, but this is often challenging because of the variety and statistical complexity of sounds that are encountered in daily life.

We first compared the performance of the IC Adaptation model and standard LN model using a correlation coefficient (CC_{raw}). To accurately assess performance, we measured predictions using a 10-fold testing procedure: for each stimulus type, we selected 10 nonoverlapping subsets of the data to be our test dataset. For each test set, we fitted an LN model using the rest of the data and used the LN model to predict responses to the test set. We measured the mean CC_{raw} (over all 10 test sets) between the LN model predictions and the actual neural responses. Using this measure suggests that there is an advantage for the IC Adaptation model over the LN model (Fig. 3A).

However, CC_{raw} is affected by neuronal response variability as well as by model performance, as can be seen from the relationship between the colors of the points in Figure 3A and the model performance. The red points show data from units with a high noise ratio; these have highly variable responses and consequently have low values of CC_{raw} . The blue points show data from

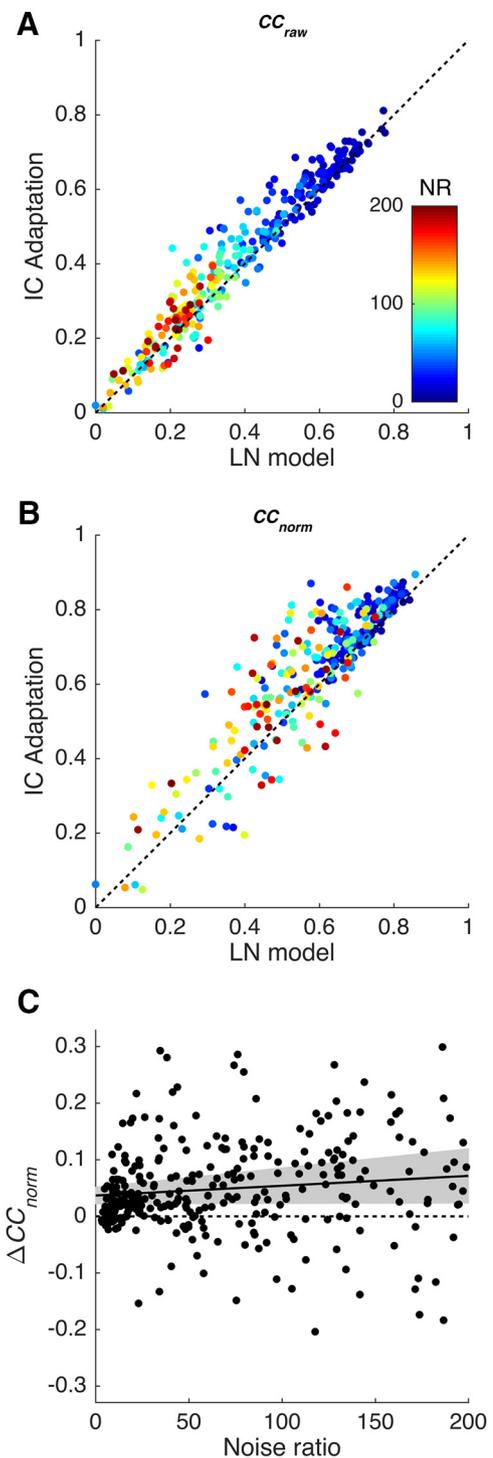


Figure 3. Comparison of the ability of the standard LN model and the IC Adaptation model to predict neural responses to natural sounds. **A, B**, Scatterplots showing the correlation coefficients between model predictions and actual neural responses (BigNat dataset). The x-axis shows performance of the standard LN model, the y-axis shows performance of the IC Adaptation model, and colors indicate the NR of each unit. **A**, Raw correlation coefficient CC_{raw} . **B**, Normalized correlation coefficient CC_{norm} . **C**, Scatterplot showing how the difference in CC_{norm} between the two models varies with NR. The solid line is a linear regression, and the shaded area shows the 95% confidence intervals on the regression.

units with a low noise ratio; these have relatively reliable responses and so have high values of CC_{raw} . To reduce this confound between model performance and neuronal response variability, we used the normalized correlation coefficient

(CC_{norm} ; see Materials and Methods) as our primary measure of model performance. CC_{norm} is the ratio of CC_{raw} to the estimated maximum possible correlation coefficient given the level of response variability in the data, CC_{max} . Since CC_{max} is a constant for each unit, using CC_{max} does not affect the relative performance of two models for that unit (i.e., for models 1 and 2, $CC_{norm}^{(1)}/CC_{norm}^{(2)} = CC_{raw}^{(1)}/CC_{raw}^{(2)}$). However, it gives a more accurate picture of the performance of the models across the whole dataset.

For both the CC_{norm} and the CC_{raw} measures, the IC Adaptation model provided better predictions than the LN model in 77% of neurons (Fig. 3A, B). The mean CC_{norm} for the IC Adaptation model was 0.64 compared with 0.59 for the LN model. This improvement in performance is highly significant ($p \ll 0.0001$, paired t test; $df = 299$).

It is conceivable that the advantage of the IC Adaptation model could, at least in part, be an artifact of data quality. For example, half-wave rectification of the stimulus spectrogram removes parts of the sound whose level is lower than the mean. This reduces the effective dimensionality of the stimulus set and may also reduce the effective number of STRF parameters that must be estimated. Because simple models require less data to constrain them than complex models, it is possible that this might give the IC Adaptation model an artificial advantage over the LN model for noisy neurons.

To rule out this possibility, we investigated the relationship between CC_{norm} and the NR. The NR quantifies the relative contributions of unpredictable and stimulus-driven variability in the neuronal responses (see Materials and Methods); a low NR indicates that a neuron was reliably driven by the stimulus. A scatterplot of the difference in CC_{norm} for the two models (Fig. 3C) shows that the advantage of the IC Adaptation model is only weakly dependent on the NR, indicating that the IC Adaptation model is generally superior to the LN model, regardless of the NR.

The IC Adaptation model also outperforms conventional LN models when tested with commonly used synthetic stimuli

An important aspect of any model is its generality. If the IC Adaptation model is a better model of cortical neurons than the standard LN model, it should provide better predictions of cortical responses to a wide range of stimuli. In one sense, testing the model on natural sounds is a good test of generality, because an appropriate collection of natural sounds will sample the space of ecologically relevant stimuli. However, it is also important to test the model using synthetic stimuli that have been widely used in neurophysiology experiments because they have been designed to exhibit well defined statistics that may provide particularly stringent tests of the model.

We therefore tested the IC Adaptation model on a second dataset (Comparison) in which several stimulus classes were presented to each unit, randomly interleaved. The classes were DRCs, TORCs, modulated noise, and natural sounds (see Materials and Methods for details). We fitted the IC Adaptation model and the LN model to each stimulus class in turn and measured predictions using CC_{norm} for reserved test data from the same class (see Materials and Methods). For every stimulus class, we found that the IC Adaptation model performs better than the LN model (Fig. 4A). This difference is significant at $p < 0.0001$ or better (paired t test) for every stimulus class except TORCs (for which $p = 0.07$). Across all stimulus classes, the mean CC_{norm} for the IC Adaptation model was 0.53 compared with 0.47 for the LN model.

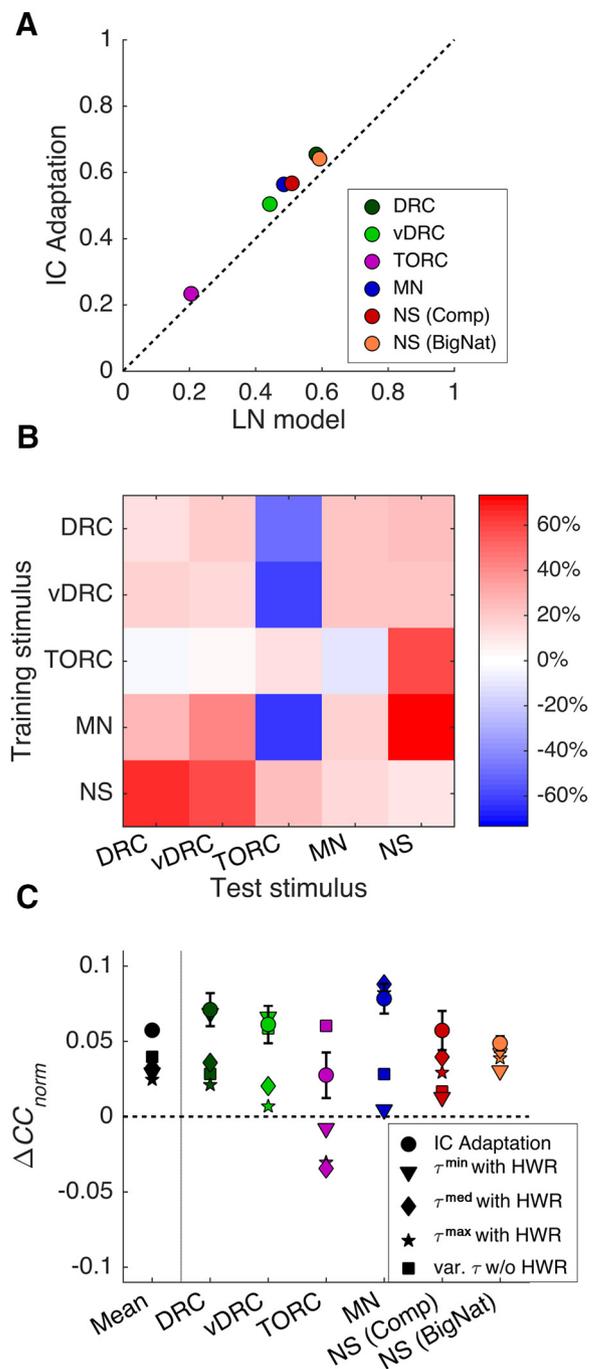


Figure 4. A, Comparison of the LN and IC Adaptation models for several stimulus classes, when models are trained and tested on the same stimulus class (dots show the mean of the within-class predictions for all units). B, Percentage improvement in mean model performance between the LN and IC Adaptation models, ΔCC_{norm} , when models are trained (rows) on one stimulus class and tested (columns) on another (cross-class predictions; Comparison dataset only). C, Difference between prediction performance of control models with fixed time constants (τ^{med} , τ^{min} , and τ^{max}) and without half-wave rectification, compared with the LN model for each stimulus class (colors as in Fig. 4A).

How well do models fare when fitted with one class of sound stimuli and tested with another?

Another important test of generality is that a model should be able to generate accurate predictions across stimulus classes. For example, if the model is trained on DRCs, it should be able to generate accurate predictions of responses to natural sounds. This has been a problem for STRF models of sensory neurons,

which often perform much worse for cross-class prediction than for within-class prediction (Olshausen and Field, 2005).

To test cross-class predictions, we fitted the IC Adaptation model and the LN model to each stimulus type in the Comparison dataset in turn and measured predictions using CC_{norm} for reserved test sets of data from other sound classes. We tested all combinations of within- and cross-class predictions by training and testing on every combination of stimulus classes and plotted the percentage difference in mean performance between the IC Adaptation model and the LN model, $\overline{\Delta CC_{\text{norm}}} = 100 \times (\overline{CC_{\text{norm}}^{\text{IC}}} - \overline{CC_{\text{norm}}^{\text{LN}}}) / \overline{CC_{\text{norm}}^{\text{LN}}}$ (Fig. 4B). In 20 of 25 cases, the IC Adaptation model performed better than the LN model. The five cases where the LN model performed better than the IC Adaptation model all involve TORC stimuli. It appears that TORCs are an unusual case where the IC Adaptation model performs particularly poorly. This may be explained by the fact that the TORCs are regularly interleaved with periods of silence. Our log-spectrograms thresholded all stimuli (see Materials and Methods) to prevent model predictions from being disproportionately affected by large negative sound levels. Normally, this has little effect on the IC Adaptation model, but during periods of silence, the IC Adaptation will adapt to the threshold value. As a result, the precise value of this threshold will significantly affect model predictions.

When trained on synthetic stimuli and tested on natural sounds (NS, right-hand column), the IC Adaptation model always outperformed the LN model, suggesting that the IC Adaptation model provides superior generalization from synthetic to natural stimuli.

It is also notable that the percentage improvements are slightly higher (though not significantly so) for cross-class predictions (off-diagonal elements; median = 22%) than for within-class predictions (main diagonal; mean = 13%; difference not significant according to a rank-sum test). This indicates that the IC Adaptation model has no negative effect on the generality of STRF models; in fact, it may improve cross-class generalization relative to the LN model.

Both frequency-dependent time constants and half-wave rectification contribute to the success of the IC adaptation model

The IC Adaptation model adds two main components to the standard LN model: high-pass modulation filtering by subtraction using exponential filters with time constants derived from Dean et al. (2008) and half-wave rectification of the resulting filtered log-spectrogram. To test whether both of these components are essential to the model, or whether either component on its own is sufficient, we investigated the effects of manipulating the IC Adaptation model in two ways.

In one set of manipulations, we produced control models where the frequency dependence of the time constants, τ_f , was removed. We replaced the frequency-dependent time constants with three fixed time constants: $\tau^{\text{med}} = 160$ ms, $\tau^{\text{min}} = 27$ ms, and $\tau^{\text{max}} = 217$ ms (corresponding to the median, minimum, and maximum of the time constants observed by Dean et al. (2008) across the range of frequencies in our log-spectrogram). In the second set, we kept the frequency-dependent time constants but removed the half-wave rectification from the IC Adaptation model. We compared the performance of these control models with the LN and IC Adaptation models for within-class predictions on all stimulus types (Fig. 4C). All of the control models perform better than the LN model when performance is

averaged across stimulus type, and the IC Adaptation model outperforms all of the controls (Fig. 4C, left-hand column).

For each individual stimulus class (Fig. 4C, other columns), the IC Adaptation model performs better than or equivalently to the controls in almost all cases. The only exception is the model without half-wave rectification (no-HWR), which performs better than the IC Adaptation model for a single stimulus class, TORCs. However, the no-HWR model performs significantly worse for DRCs, modulated noise, and natural sounds (Comparison dataset). Overall, these results suggest that both frequency dependence of the time constants and half-wave rectification are important components of the IC Adaptation model.

Why are the adaptation time constants in IC distributed as they are?

To investigate why the time constants of adaptation in the IC exhibit the characteristic frequency-dependent distribution that has been described previously, we asked whether sound levels in natural sounds exhibit a similar frequency-dependent distribution, which would imply that the IC time constants are optimized to match the statistical structure of natural sounds.

Specifically, we assumed that mean sound level is a nuisance variable and that the role of adaptation to mean sound level in the IC is to subtract this nuisance variable from the neural representation of sound. To perform this subtraction, the IC must estimate the mean level in each frequency band. If the mean level is stable over time, an adaptation process with a short time constant should be able to reliably estimate the mean. If the mean level is unstable, a longer time constant will be required to produce a reliable estimate. We therefore measured the stability of the mean level using the autocorrelation of the spectrogram of natural sounds. A narrow autocorrelation function indicates that sound level is poorly correlated over time (unstable), whereas a wide autocorrelation function indicates that sound level is well correlated over time (stable). Given the results of Dean et al. (2008), we might expect that low frequencies will have narrower autocorrelation functions than high frequencies.

For a large set of natural sounds, we took the spectrogram of each sound and measured the autocorrelation of the spectrogram for each frequency band (see Materials and Methods). The mean (across all sounds) of the resulting autocorrelation functions is shown in Figure 5A. The width of the autocorrelation function (Fig. 5, black lines) varies with frequency, as predicted from the data of Dean et al. (2008). The widths range from 10 ms at the lowest frequencies (500 Hz) to 260 ms at the highest frequencies (32 kHz).

We also investigated what time constants are required to optimally estimate the mean sound level in different frequency bands. We took a large set of natural sounds and divided them into seven broad categories (see Materials and Methods). For each category, we estimated a set of linear kernels (one for each frequency channel) that optimally predict the mean level of the next T_{av} ms of sound, based on the past 2.5 s of sound. Each kernel was constrained to have an exponential shape. The time constants of the exponential kernels were optimized separately for each frequency channel, and for 8 log-spaced values of T_{av} , (between 15 ms and 1000 ms).

The time constants for the lowest and the highest frequency bands (centered at 70 Hz and 20 kHz, respectively) are plotted as a function of T_{av} in Figure 5B. In all cases, the mean time constants for low frequencies (Fig. 5B, red line; error bars show SE) are consistently higher than those for high frequencies (Fig. 5B, blue line). This confirms that longer time constants are required

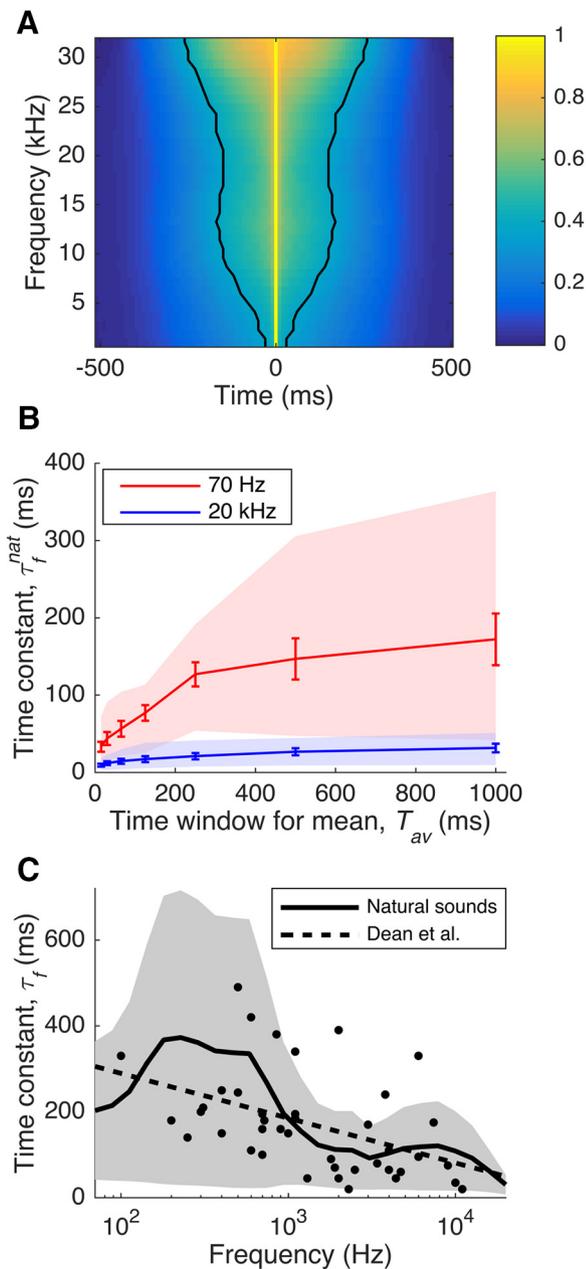


Figure 5. *A*, Autocorrelation of the spectrogram of an ensemble of natural sounds. The black lines indicate the point where the autocorrelation has decreased to 1/e. *B*, Optimal time constants for predicting the mean sound level of natural sounds in a window of length T_{av} ms into the future. This is plotted as a function of the window size, for frequency bands centered on a low value (70 Hz; red line) and a high value (20 kHz; blue line). Lines indicate mean, error bars show SEM, and shaded regions indicate entire range. *C*, Optimal time constants for prediction of the mean level of natural stimuli in each frequency band. The shaded region indicates the range of time constants in that frequency band; the solid line indicates the middle of the range. The dots indicate time constants observed in guinea pig IC by Dean et al. (2008), and the dashed line is the linear regression of those time constants against log(frequency).

to reliably estimate the mean for low frequencies. We find that as the length of the averaging window, T_{av} , increases, the optimal time constants increase. This effect begins to saturate above 250 ms. We also find that there is substantial variation between sound categories (the range from the minimum to maximum value is shown by the red and blue shaded regions). However, for every sound category, the optimal time constants are typically larger for low-frequency sounds than for high-frequency sounds, suggesting that this is a consistent property of natural sounds.

To compare the time constants measured by Dean et al. (2008) with the time constant range that is optimal for natural sounds, we took the results of the above analysis and measured the whole range of time constants obtained (across all sound categories and values of T_{av} from 250 ms upward) for each frequency band. The results, plotted in Figure 5C (solid line indicates center; shaded region indicates range), are in good agreement with the time constants for IC neurons from Dean et al. (2008) (dots and regression line). The two datasets show very similar inverse relationships between time constant and frequency. Most of the IC data fall within the range of time constants that might be considered optimal. Nevertheless, there is significant variation, both within the data of Dean et al. (2008) and among the time constants for different sound categories. This may reflect optimization of subpopulations of neurons in IC for different subsets of natural sounds with different temporal autocorrelations. Overall, these results support our hypothesis that the time constants of adaptation to mean level in the IC are indeed optimized to enable neurons in the auditory midbrain to estimate and then subtract the mean sound level in each frequency band.

Discussion

Developing predictive, quantitative models of the response properties of individual neurons is fundamental to our ability to describe and understand information processing in the brain. Although STRF-based models have their limitations, they remain valuable and provide simple models for describing the behavior of sensory neurons. Many of the more sophisticated models (Sharpee et al., 2004, 2011; Ahrens et al., 2008; Atencio et al., 2009; Calabrese et al., 2011; Schinkel-Bielefeld et al., 2012) require many more parameters and may also be difficult to interpret biologically. Thus, a key challenge is to extend STRF models so that they more accurately describe neuronal behavior, while remaining simple and biologically relevant. Here we have shown that we can significantly improve STRF models of neurons in the auditory cortex by introducing a simple nonlinear input transform that reflects adaptation to stimulus statistics in the midbrain.

Advantages and power of nonlinear input stages

It is widely accepted that it can be useful to add a nonlinear output stage to the basic STRF model, resulting in an LN model (Chichilnisky, 2001). A nonlinear output stage has numerous advantages over the purely linear model. Sigmoid (or similar) functions can model threshold and saturation effects that are present in real neurons. However, the output of an LN model is only a simply transformed version of the output of the linear model. Introducing a nonlinear input stage is potentially far more powerful, because it allows the model to perform potentially quite complex nonlinear computations. In principle, unlimited nonlinear processing of the input can be performed before the linear summation stage, so that the full NLN model can perform complex computations. However, in practice it is difficult to harness this power because models that involve complex input transformations typically involve large numbers of free parameters, which are difficult to estimate given that neurophysiological datasets are always limited and noisy. We therefore need a way to introduce appropriate nonlinear transformations of the input without introducing many free parameters.

Here we have circumvented this problem by constraining the model nonlinear input transformations to replicate known operations of early stages of the sensory pathway. To the extent that the sensory systems are hierarchically organized, we can charac-

terize the input by recording the response properties of neurons at earlier stages of processing. In the present case, we have used a model of the characteristics of neurons in the IC as the input to a model of the behavior of cortical neurons. Because the parameters of this input transformation can be determined by recording neuronal responses in the IC, it is possible to incorporate a well characterized transformation without introducing any free parameters whatsoever, which is what we have done here.

Relationship with other models

The IC Adaptation model includes a nonlinear input transformation with three components: adaptation to the stimulus mean, half-wave rectification, and frequency-dependent time constants of adaptation. Other models in the auditory literature have introduced nonlinear input stages to STRF models but have not examined this particular combination. For example, the synaptic depression model by David and colleagues (David et al., 2009; David and Shamma, 2013) contains similar adaptation to the stimulus level, but without half-wave rectification or frequency-dependent time constants. It is therefore similar to one of the controls used here. We find that both the half-wave rectification and the frequency-dependent time constants introduced by the IC Adaptation model significantly improve the power of the model to predict neural responses to most new stimuli.

The context model described by Ahrens et al. (2008) has a considerably greater model complexity than our IC Adaptation model and is therefore, in principle, capable of far more powerful nonlinear transformations, including nonmonotonicity. However, it also introduces many additional free parameters, which are difficult to fit reliably to neural data. This limits the improvements in predictive power that can be achieved in practice. The IC Adaptation model, in contrast, has no more free parameters than the LN model and therefore provides some of the benefits of the context model without introducing extra model complexity.

Time constants of adaptation

This study builds on the results of Dean et al. (2008), who found that neurons in the guinea pig IC adapt to the mean level of recent sound stimulation and that this adaptation has frequency-dependent time constants. We found that the specific time constants measured in that study were valuable in improving our models of ferret cortical neurons.

We show (Fig. 5C) that the time constants are optimized for a specific representation of the sound waveform. We assumed that the time-varying mean sound level in each frequency band, μ_{tf} , is a nuisance variable, which does not need to be included in the neuronal representation of sound. If this is the case, then a plausible role for adaptation to mean sound level in the IC is to subtract an estimate, $\hat{\mu}_{tf}$, of the time-varying mean (in a time window, T_{av}) from the sound level, X_{tf} , so that the responses of IC neurons are functions of $X_{tf} - \hat{\mu}_{tf}$ rather than μ_{tf} itself. We also assumed that the adaptation process estimates future values of $\hat{\mu}_{tf}$ by exponential averaging over recent values of X_{tf} . Finally, we assumed that the time constants of this adaptation process are optimized so that, for each frequency band, $\hat{\mu}_{tf}$ is as close as possible to the true mean, μ_{tf} . Using only these assumptions, we were able to derive optimal time constants for each frequency band in the spectrogram and found that these time constants are a good match for the real time constants measured in IC. This supports our hypothesis that adaptation in IC is optimized to subtract the time-varying mean in each frequency band of natural sounds.

It may initially seem surprising that time constants measured for neurons in the guinea pig IC should be relevant for a different species. However, since our natural sound analysis makes no assumptions that are specific to guinea pigs, the time constants should be similar for a range of species. While our assumption that mean sound level is a nuisance variable is a good general principle, we expect that there is some behaviorally valuable information in the absolute sound level that will also need to be transmitted. It is notable that in the auditory system adaptation is not complete, and even neurons which adapt optimally will not perfectly estimate and subtract μ_{tf} . As a result, some residual information about mean level will still be transmitted to the auditory cortex.

Future directions

The IC Adaptation model is a simple, easily implemented extension of classical STRF models of auditory neurons. It improves predictions of the behavior of neurons in the auditory cortex by incorporating a model of adaptation to stimulus statistics at an earlier stage of processing. Because it introduces no free parameters, it neither increases the complexity of the model nor the amount of data required to fit it.

In the present study, we have modeled adaptation to mean sound level in the IC and applied this to the responses of neurons in A1/AAF. It is likely that this work can be generalized to other structures in the auditory system. For example, further adaptation to stimulus mean is present in the responses of cortical neurons (Rabinowitz et al., 2013) and may have a thalamic or cortical origin. Future studies could experimentally characterize adaptation properties in these structures and use this to improve models of neurons in primary and higher auditory cortices.

In the visual system, adaptation to stimulus mean luminance is also present (Dawis et al., 1984; Rodieck, 1998; Mante et al., 2005) but is not yet routinely incorporated into receptive-field models of visual neurons (for review, see Sharpee, 2013). Nishimoto and Gallant (2011) found that incorporating luminance (and contrast) normalization did not significantly improve models of processing in MT; however, their model of normalization was a general one that did not closely match the characteristics of any particular adaptation process, and their stimuli contained a relatively narrow range of luminances. An approach similar to the one used here, using measured time constants of adaptation, may be better able to improve predictions of neural responses to motion and other visual stimuli.

Finally, neurons at multiple levels of the visual (Fairhall et al., 2001; Carandini and Heeger, 2012), somatosensory (Garcia-Lazaro et al., 2007), and auditory (Rabinowitz et al., 2011) systems show adaptation to higher-order statistics such as stimulus variance. Using an approach very similar to the one presented here, it should be possible to incorporate adaptation to higher-order stimulus statistics into neural models at multiple levels of different sensory pathways.

References

- Aertsen AM, Johannesma PI (1981) The spectro-temporal receptive field. *Biol Cybern* 42:133–143. [CrossRef Medline](#)
- Aertsen AMHJ, Johannesma PIM, Hermes DJ (1980) Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biol Cybern* 38:235–248. [CrossRef](#)
- Ahrens MB, Linden JF, Sahani M (2008) Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectro-temporal methods. *J Neurosci* 28:1929–1942. [CrossRef Medline](#)
- Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58:956–966. [CrossRef Medline](#)

- Atencio CA, Sharpee TO, Schreiner CE (2009) Hierarchical computation in the canonical auditory cortical circuit. *Proc Natl Acad Sci U S A* 106:21894–21899. [CrossRef Medline](#)
- Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW (2009) Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J Neurosci* 29:2064–2075. [CrossRef Medline](#)
- Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS One* 6:e16104. [CrossRef Medline](#)
- Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62. [CrossRef Medline](#)
- Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network* 12:199–213. [CrossRef Medline](#)
- Christianson GB, Sahani M, Linden JF (2008) The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J Neurosci* 28:446–455. [CrossRef Medline](#)
- Dahmen JC, Keating P, Nodal FR, Schulz AL, King AJ (2010) Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron* 66:937–948. [CrossRef Medline](#)
- David SV, Shamma SA (2013) Integration over multiple timescales in primary auditory cortex. *J Neurosci* 33:19154–19166. [CrossRef Medline](#)
- David SV, Mesgarani N, Fritz JB, Shamma SA (2009) Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *J Neurosci* 29:3374–3386. [CrossRef Medline](#)
- Dawis S, Shapley R, Kaplan E, Tranchina D (1984) The receptive field organization of X-cells in the cat: spatiotemporal coupling and asymmetry. *Vision Res* 24:549–564. [CrossRef Medline](#)
- Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8:1684–1689. [CrossRef Medline](#)
- Dean I, Robinson BL, Harper NS, McAlpine D (2008) Rapid neural adaptation to sound level statistics. *J Neurosci* 28:6430–6438. [CrossRef Medline](#)
- deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443. [CrossRef Medline](#)
- Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J Neurosci* 22:4114–4131. [Medline](#)
- Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412:787–792. [CrossRef Medline](#)
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223. [CrossRef Medline](#)
- Garcia-Lazaro JA, Ho SS, Nair A, Schnupp JW (2007) Shifting and scaling adaptation to dynamic stimuli in somatosensory cortex. *Eur J Neurosci* 26:2359–2368. [CrossRef Medline](#)
- Gill P, Zhang J, Woolley SM, Fremouw T, Theunissen FE (2006) Sound representation methods for spectro-temporal receptive field estimation. *J Comput Neurosci* 21:5–20. [CrossRef Medline](#)
- Gourévitch B, Noreña A, Shaw G, Eggermont JJ (2009) Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. *Cereb Cortex* 19:1448–1461. [CrossRef Medline](#)
- Hsu A, Borst A, Theunissen FE (2004) Quantifying variability in neural responses and its application for the validation of model predictions. *Network* 15:91–109. [CrossRef Medline](#)
- Joris PX, Yin TC (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. *J Acoust Soc Am* 91:215–232. [CrossRef Medline](#)
- Kadir SN, Goodman DF, Harris KD (2014) High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput* 26:2379–2394. [CrossRef Medline](#)
- Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9:85–111. [CrossRef Medline](#)
- Kvale MN, Schreiner CE (2004) Short-term adaptation of auditory receptive fields to dynamic stimuli. *J Neurophysiol* 91:604–612. [Medline](#)
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol* 90:2660–2675. [CrossRef Medline](#)
- Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24:1089–1100. [CrossRef Medline](#)
- Malone BJ, Scott BH, Semple MN (2010) Temporal codes for amplitude contrast in auditory cortex. *J Neurosci* 30:767–784. [CrossRef Medline](#)
- Mante V, Frazor RA, Bonin V, Geisler WS, Carandini M (2005) Independence of luminance and contrast in natural scenes and in the early visual system. *Nat Neurosci* 8:1690–1697. [CrossRef Medline](#)
- McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71:926–940. [CrossRef Medline](#)
- Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87:516–527. [Medline](#)
- Nagel KI, Doupe AJ (2006) Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* 51:845–859. [CrossRef Medline](#)
- Nelson PC, Carney LH (2007) Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *J Neurophysiol* 97:522–539. [CrossRef Medline](#)
- Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31:14551–14564. [CrossRef Medline](#)
- Olshausen BA, Field DJ (2005) How close are we to understanding V1? *Neural Comput* 17:1665–1699. [CrossRef Medline](#)
- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2011) Contrast gain control in auditory cortex. *Neuron* 70:1178–1191. [CrossRef Medline](#)
- Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2012) Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J Neurosci* 32:11271–11284. [CrossRef Medline](#)
- Rabinowitz NC, Willmore BD, King AJ, Schnupp JW (2013) Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol* 11:e1001710. [CrossRef Medline](#)
- Rees A, Møller AR (1983) Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear Res* 10:301–330. [CrossRef Medline](#)
- Rodiek RW (1998) *The first steps in seeing*. Sunderland, MA: Sinauer.
- Rutkowski RG, Shackleton TM, Schnupp JW, Wallace MN, Palmer AR (2002) Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the guinea pig. *Audiol Neurootol* 7:214–227. [CrossRef Medline](#)
- Sahani M, Linden JF (2003) How linear are auditory cortical responses? *Adv Neural Inform Process Syst* 15:125–132.
- Schinkel-Bielefeld N, David SV, Shamma SA, Butts DA (2012) Inferring the role of inhibition in auditory processing of complex natural stimuli. *J Neurophysiol* 107:3296–3307. [CrossRef Medline](#)
- Schnupp JW, Mrsic-Flogel TD, King AJ (2001) Linear processing of spatial cues in primary auditory cortex. *Nature* 414:200–204. [CrossRef Medline](#)
- Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16:223–250. [CrossRef Medline](#)
- Sharpee TO (2013) Computational identification of receptive fields. *Annu Rev Neurosci* 26:103–120. [CrossRef Medline](#)
- Sharpee TO, Atencio CA, Schreiner CE (2011) Hierarchical representations in the auditory cortex. *Curr Opin Neurobiol* 21:761–767. [CrossRef Medline](#)
- Simoncelli EP, Paninski L, Pillow J, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. In: *The cognitive neurosciences III* (Gazzaniga MS, ed), pp 327–338. Cambridge, MA: MIT.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331. [Medline](#)
- Touryan J, Felsen G, Dan Y (2005) Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45:781–791. [CrossRef Medline](#)
- Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. *Curr Opin Neurobiol* 17:423–429. [CrossRef Medline](#)
- Watkins PV, Barbour DL (2008) Specialized neuronal adaptation for preserving input sensitivity. *Nat Neurosci* 11:1259–1261. [CrossRef Medline](#)
- Wen B, Wang GI, Dean I, Delgutte B (2009) Dynamic range adaptation to sound level statistics in the auditory nerve. *J Neurosci* 29:13797–13808. [CrossRef Medline](#)
- Willmore B, Smyth D (2003) Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. *Network* 14:553–577. [CrossRef Medline](#)
- Willmore BD, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. *J Neurosci* 30:2102–2114. [CrossRef Medline](#)