

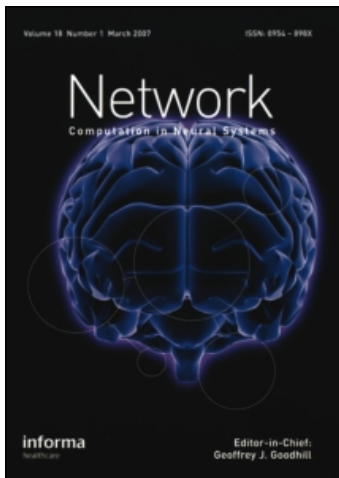
This article was downloaded by: [University Of Maryland]

On: 4 September 2008

Access details: Access Details: [subscription number 748925376]

Publisher Informa Healthcare

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Network: Computation in Neural Systems

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713663148>

Estimating sparse spectro-temporal receptive fields with natural stimuli

Stephen V. David ^a; Nima Mesgarani ^a; Shihab A. Shamma ^a

^a Institute for Systems Research, University of Maryland, College Park, MD, USA

First Published on: 07 September 2007

To cite this Article David, Stephen V., Mesgarani, Nima and Shamma, Shihab A.(2007)'Estimating sparse spectro-temporal receptive fields with natural stimuli',*Network: Computation in Neural Systems*,18:3,191 — 212

To link to this Article: DOI: 10.1080/09548980701609235

URL: <http://dx.doi.org/10.1080/09548980701609235>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Estimating sparse spectro-temporal receptive fields with natural stimuli

STEPHEN V. DAVID, NIMA MESGARANI,
& SHIHAB A. SHAMMA

Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

(Received 14 March 2007; revised 5 July 2007; accepted 2 August 2007)

Abstract

Several algorithms have been proposed to characterize the spectro-temporal tuning properties of auditory neurons during the presentation of natural stimuli. Algorithms designed to work at realistic signal-to-noise levels must make some prior assumptions about tuning in order to produce accurate fits, and these priors can introduce bias into estimates of tuning. We compare a new, computationally efficient algorithm for estimating tuning properties, *boosting*, to a more commonly used algorithm, *normalized reverse correlation*. These algorithms employ the same functional model and cost function, differing only in their priors. We use both algorithms to estimate spectro-temporal tuning properties of neurons in primary auditory cortex during the presentation of continuous human speech. Models estimated using either algorithm, have similar predictive power, although fits by boosting are slightly more accurate. More strikingly, neurons characterized with boosting appear tuned to narrower spectral bandwidths and higher temporal modulation rates than when characterized with normalized reverse correlation. These differences have little impact on responses to speech, which is spectrally broadband and modulated at low rates. However, we find that models estimated by boosting also predict responses to non-speech stimuli more accurately. These findings highlight the crucial role of priors in characterizing neuronal response properties with natural stimuli.

Keywords: *Auditory cortex, speech, reverse correlation, boosting*

Introduction

Recently, interest has grown in characterizing the response properties of sensory neurons under natural stimulus conditions (Theunissen et al. 2001; David and Gallant 2005; Wu et al. 2006). This approach is useful for understanding how well functional models developed using simple synthetic stimuli generalize to conditions encountered during natural behavior (Theunissen et al. 2000; David et al. 2004). In this study, we evaluate a new method for characterizing functional properties from their responses to natural stimuli: boosting (Friedman et al. 2000; Willmore et al. 2005; Zhang and Yu 2005; Willmore et al. 2006).

In auditory systems, the functional properties of single neurons are often described by a spectro-temporal receptive field (STRF) (Kowalski et al. 1996; Theunissen et al. 2001; Sahani and Linden 2003; Machens et al. 2004). The STRF is a linear model in the spectral domain of sound and is well-suited for systems identification methodologies. Classically, measurements of STRFs have been restricted to the domain of white noise stimuli because spike-triggered averaging, the most commonly used estimation method, produces biased estimates for non-white stimuli, including natural stimuli (Theunissen et al. 2001). More recently, a new method, normalized reverse correlation (NRC), has been developed that converges on unbiased STRF estimates for arbitrary, non-white stimuli (Theunissen et al. 2001), along with some slight variants (Willmore and Smyth 2003; Machens et al. 2004). Studies using NRC have identified large significant changes in tuning under natural stimulus conditions that are not observed during stimulation by synthetic stimuli both in auditory (Theunissen et al. 2001; Woolley et al. 2005) and visual systems (David et al. 2004).

Because natural stimuli contain strong autocorrelations, some dimensions have much lower variance than others (Theunissen et al. 2001). Signal-to-noise levels for response properties along the low-variance dimensions tend to be poor. Algorithms for STRF estimation such as NRC require regularization to prevent overfitting to noise along the low-variance dimensions. Regularization introduces a prior that imposes a constraint on the STRF estimate, independent of the data to be fit. Under noisy conditions, a prior can bias STRF estimates, and it is critical to account for the effects of this bias when drawing conclusions about tuning (David et al. 2004; Woolley et al. 2005).

Since the development of NRC, several other methods have been developed for estimating receptive field properties using natural stimuli (Ringach et al. 2002; Willmore and Smyth 2003; Prenger et al. 2004; Sharpee et al. 2004; Touryan et al. 2005; Willmore et al. 2005; Rapela et al. 2006). These algorithms are largely variants of gradient descent techniques, in which STRF estimates are repeatedly adjusted by small amounts in the direction that minimizes fit error. Gradient descent algorithms can be used to fit data with the same functional model (linear STRF) and cost function (minimum mean-squared error) as NRC; thus they converge on the same theoretical solution as NRC. However, each algorithm uses a different prior than NRC, which can bias STRF estimates for finite noisy data (Wu et al. 2006).

One new method for STRF estimation, boosting, is appealing because it requires a relatively light computational load and provides easily interpretable results (Friedman et al. 2000; Zhang and Yu 2005). This method has been shown to be effective for characterizing spatiotemporal responses properties in extrastriate visual

cortex (Willmore et al. 2005, 2006). However, a direct comparison has not yet been made between boosting and other STRF estimation methods.

In this study, we compare STRFs estimated by boosting and NRC under natural stimulus conditions. We estimate STRFs from the responses of isolated neurons in primary auditory cortex (A1) to continuous human speech (Garfalo 1988). Speech, while not a complete sampling of all possible sounds, is a complex mammalian vocalization that shares high-order statistical properties with a broad class of natural sounds (Smith and Lewicki 2006). We compare boosted and NRC STRFs in terms of their predictive power and their basic tuning properties. We find that boosted STRFs predict responses to speech and synthetic non-speech stimuli significantly better than NRC STRFs. The differences in predictive power result directly from differences in the tuning bandwidth of STRFs estimated by the two methods. These results suggest that boosting provides a more accurate characterization of spectro-temporal tuning under these experimental conditions.

Methods

Experimental procedures

Auditory response properties were recorded from single neurons in primary auditory cortex (A1) of five awake passive ferrets. All experimental procedures were conformed to standards specified by the National Institutes of Health and the University of Maryland Animal Care and Use Committee.

Surgical preparation. Animals were implanted with a stainless steel head post to allow for stable recording. While under anesthesia (nembutol and halothane), the skin and muscles on the top of the head were retracted from the central 4 cm diameter of skull. Several titanium bone screws were attached to the skull and the head post was glued on the midline. The entire site was then covered with dental acrylic and the skin around the implant was allowed to heal. Analgesics (banamine) and antibiotics (baytril) were administered under veterinary supervision until recovery.

After recovery from surgery, animals were acclimatized to a custom restrainer where their heads were held fixed. A small craniotomy (1 mm diameter) was then drilled through the acrylic and skull over A1. The craniotomy was cleaned daily to prevent infection. After recordings were completed in one hemisphere, the site was closed with a layer of dental acrylic, and the same procedure was repeated in the other hemisphere.

Neurophysiology. Single unit activity was recorded using tungsten micro-electrodes (1–5 MO, FHC). For three animals a single electrode was used and its position was controlled using a pneumatic microdrive (Narishige). Electrical activity was amplified with a custom amplifier, converted to a digital signal (National Instruments), and recorded using a custom data acquisition system. For the remaining two animals, one to four electrodes were controlled by independent microdrives and activity was recorded using a commercial data acquisition system

(Alpha-Omega). For all experiments, spike events were extracted from the raw electrophysiological recordings using custom template matching software. Only single units with clear (>95%) isolation were included in the analysis data set.

Stimuli were presented from digital recordings using custom software. The digital signals were transformed to analog (National Instruments), equalized to achieve flat gain (Rane), and amplified (Radio Shack) to the desired sound level. These signals were presented through a small speaker (Etymotics) inserted in the ear contralateral to the recording site. Before each recording, the equalizer was calibrated to the acoustical properties of the speaker insertion.

Upon identification of a recording site with isolatable units, a sequence of random tones (100 ms duration, 500 ms separation) was used to measure latency and spectral tuning. Neurons were verified as being in A1 by their tonotopic organization, latency and simple frequency tuning curves (Kowalski et al. 1996; Bizley et al. 2005).

Speech stimuli. Continuous speech stimuli were drawn from a standard speech database (TIMIT; Garfalo 1988). Stimuli were sentences (3–4 sec), each sampled from a different speaker (balanced across gender). For each neuron, 30–90 different sentences were presented at 60–65 dB SPL for 5 repetitions. The original stimuli were recorded at 16 kHz but were upsampled to 44 kHz before presentation.

Temporally orthogonal ripple stimuli. In order to test how well STRFs generalize across stimulus classes, responses were also recorded to temporally orthogonal ripple combinations (TORCs; Klein et al. 2000). TORCs provide an efficient probe of spectro-temporal response properties by approximating white noise in the spectral domain. The set of 30 TORCs probed tuning over 5 octaves (250–8000 Hz or 500–16 000 Hz, chosen to span the range of frequencies at which pure tones drive responses), with a spectral resolution of 1.2 cycles per octave and temporal envelope resolution of 24 Hz or 48 Hz (5 repetitions, 3 s duration, 44 kHz sampling).

Spectro-temporal receptive field estimation

Linear spectro-temporal model. Neurons in A1 are tuned to stimulus frequency but are rarely phase-locked to oscillations of the sound waveform (Kowalski et al. 1996; Bizley et al. 2005). To describe tuning properties of such neurons, it is useful to represent auditory stimuli in terms of their spectrogram. The spectrogram of a sound waveform transforms the stimulus into a time-varying function of energy in each frequency band (Figure 1). This transformation removes the phase of the carrier signal so that the mapping from the spectrogram to neuronal firing rate can be described by a linear function.

For a stimulus spectrogram $s(x, t)$ and instantaneous neuronal firing rate $r(t)$ sampled at times $t = 1, \dots, T$, the linear STRF is defined as the linear mapping

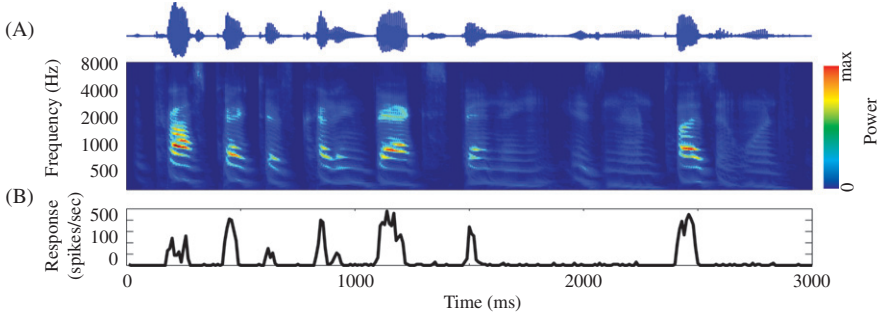


Figure 1. (A) Continuous speech stimulus presented to neurons in primary auditory cortex (A1). To estimate the STRF, the raw stimulus waveform (top) was filtered by a bank of logarithmically spaced gammatone filters that simulate the representation of stimuli in the auditory nerve. Filter center frequencies ranged from 100 to 8000 Hz. (B) The peristimulus time histogram response of a single neuron to five repeated presentations of the stimulus in (A), 10 ms bins. This neuron shows clear responses to syllables with power near 700 Hz.

(Kowalski et al. 1996; Klein et al. 2000; Theunissen et al. 2001),

$$r(t) = \sum_{\substack{x=1, \\ u=0}}^{X,U} h(x, u)s(x, t - u) + e(t) \tag{1}$$

Each coefficient of h indicates the gain applied to frequency channel x at time lag u . Positive values indicate components of the stimulus correlated with increased firing, and negative values indicate components correlated with decreased firing. The residual, $e(t)$, represents components of the response (nonlinearities and noise) that cannot be predicted by the linear model.

To simplify the description of estimation algorithms used in this study, we rewrite the STRF in a linear algebraic form. First, we define the STRF as a vector, h , by concatenating the STRF gain at each frequency and time lag into a single vector of length $Y = XU$,

$$h = \begin{bmatrix} h(1, 1) \\ h(2, 1) \\ \vdots \\ h(1, 2) \\ \vdots \\ h(X, U) \end{bmatrix} \tag{2}$$

We then define a $T \times Y$ stimulus matrix, S , such that each row, t , contains the stimulus coefficients at time t and the preceding u time bins,

$$S = \begin{bmatrix} s(1, 1) & s(2, 1) & \cdots & s(1, 0) & \cdots & s(X, 1 - U) \\ s(1, 2) & s(2, 2) & & s(1, 1) & & s(X, 2 - U) \\ \vdots & & & & \ddots & \vdots \\ s(1, T) & s(2, T) & \cdots & s(1, T - 1) & \cdots & s(X, T - U) \end{bmatrix} \tag{3}$$

The response, a vector of length T , is simply the product of the stimulus and STRF, plus a vector of residuals, e ,

$$r = Sh + e \quad (4)$$

STRF estimation by boosting. STRFs were estimated from the responses to the speech stimuli by boosting (Zhang and Yu 2005). Boosting converges on an unbiased estimate of the linear mapping between stimulus and response, regardless of autocorrelation in the stimulus. Several versions of boosting algorithms exist that can be used to estimate STRFs. In this study, we used forward stagewise fitting, which employs a simple iterative algorithm (Friedman et al. 2000).

Initially, the STRF is set to zero, $h_0 = 0$. During each iteration, i , the mean-squared error is calculated for the prediction after incrementing or decrementing each STRF parameter by a small amount, ε . All possible increments are specified by two parameters, an index, $\gamma = 1, \dots, Y$, and a sign, $\zeta = -1$ or 1 , such that,

$$\begin{aligned} \Delta h_{\gamma, \zeta}(y) &= \zeta \varepsilon, & y = \gamma \\ &= 0, & \text{otherwise} \end{aligned} \quad (5)$$

The best increment is the one that produces the largest decrease in the mean-squared error,

$$(y_i, z_i) = \arg \min_{\gamma, \zeta} \sum_{t=1}^T (r - S(h_{i-1} + \Delta h_{\gamma, \zeta}))^2 \quad (6)$$

The increment is added to the STRF,

$$h_i = h_{i-1} + \Delta h_{y_i, z_i} \quad (7)$$

and this procedure is repeated until an additional increment/decrement only adds noise to the STRF estimate.

Implementing boosting requires two *hyperparameters*, i.e., parameters that affect the final STRF estimate but that are not contained explicitly in the stimulus/response data. These are (1) the step size, ε , and (2) the number of iterations to complete before stopping. Generally, step size can be made arbitrarily small for the best fits. Extremely small step sizes are computationally inefficient, as they require many iterations to converge. We fixed ε to be a small fraction of the ratio between stimulus and response variance,

$$\varepsilon = \frac{1}{50} \sqrt{\frac{\text{var}(r(t))}{\text{var}(s(x, t))}} \quad (8)$$

Here, stimulus variance is averaged across all spectral channels. Despite differences in variance across spectral channels, this heuristic produced accurate estimates and required relatively little computation time. Increasing or decreasing ε by a factor of two had no effect on STRFs. Of course, different values of ε are likely to be optimal for different data sets.

To optimize the second hyperparameter, the number of iterations to complete before stopping, we used cross validation and early stopping. We reserved a small part (5%) of the fit data from the main boosting procedure. After each iteration, we tested the ability of the STRF to predict responses in the reserved set. The optimal

stopping point was determined as the iteration when the STRF failed to improve predictions. Note that this reserved set was distinct from the validation data used finally to evaluate each STRF (see further).

STRF estimation by NRC. We also estimated STRFs from the same data using NRC, a variant of classical linear regression that has been used to estimate STRFs from natural stimuli in the visual and auditory systems (Theunissen et al. 2000; David et al. 2004). Like boosting, NRC fits a linear STRF that minimizes the mean-squared error between predicted and observed neuronal response. A detailed description of a computationally efficient NRC algorithm is described in (Theunissen et al. 2001). Here, we provide a brief description of NRC for comparison to boosting.

In a system with no noise, the minimum mean-squared error estimate of the STRF for response vector, r , and stimulus matrix, S , is,

$$h_{\text{ideal}} = \frac{1}{T} C_{ss}^{-1} S^T r \tag{9}$$

The matrix C_{ss}^{-1} is the inverse of the $Y \times Y$ stimulus autocorrelation matrix, $C_{ss} = S^T S / T$, and the superscript T indicates the transpose operation. This estimate is theoretically optimal, but, in practice, estimating the STRF by Equation 9 for stimuli with strong autocorrelations can amplify noise excessively (Figure 6B). In a correlated stimulus, some dimensions have higher variance than others. Low variance dimensions provide very little modulating energy and thus make it difficult to measure correlations between those dimensions and the response. The inverse autocorrelation matrix normalizes the variance along each dimension to be the same. When variance is low, normalization requires dividing by a small number and often amplifies noise in parameter estimates. To minimize these effects, NRC uses a *pseudoinverse* to approximate the inverse of the stimulus autocorrelation matrix. Dimensions that are below some noise threshold are forced to be zero.

To compute the pseudoinverse, a singular value decomposition is applied to the autocorrelation matrix,

$$C_{ss} = U^T \Lambda U \tag{10}$$

The columns of U contain the unit-norm eigenvectors of C_{ss} , and the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_Y)$ contains the corresponding eigenvalues ordered from largest to smallest. A tolerance value, τ , specifies the fraction of stimulus variance to preserve in the pseudoinverse. The number of stimulus dimensions to preserve, m , is computed,

$$m = \arg \max \frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_m + \dots + \lambda_Y} < \tau \tag{11}$$

The pseudoinverse, C_{approx}^{-1} , is then computed,

$$C_{\text{approx}}^{-1} = U \Lambda_{\text{approx}}^{-1} U^T = U \text{diag} \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_m}, 0, \dots, 0 \right) U^T \tag{12}$$

And the final NRC estimate of the STRF is,

$$h_{\text{NRC}} = \frac{1}{T} C_{\text{approx}}^{-1} S^T r \quad (13)$$

Implementing NRC requires the selection of a single hyperparameter, the tolerance value, τ . To choose τ , we used a cross-validation procedure similar to the procedure for selecting the number of iterations in the boosting algorithm. A small subset was reserved from the estimation data before using Equation 13 to compute h_{NRC} . The STRF was estimated for a range of tolerance values, $\tau = 0.9 \dots 0.999999$ (i.e., $1 - 10^{-1} \dots 1 - 10^{-6}$, Figure 5B). Each STRF was then used to predict responses in the reserved data and the tolerance value that produced the best prediction was used for the final STRF estimate. As in the case of boosted STRFs, the final STRF was then validated with a separate data set that was not used at any stage of estimation (see subsequently).

Data preprocessing. The same preprocessing was applied to the speech data prior to STRF estimation by boosting and NRC. Spectrograms were generated from the stimulus sound pressure waveforms using a 128-channel rectifying filter bank that simulated processing by the auditory nerve (Yang et al. 1992). Filters had center frequencies ranging from 100 to 8000 Hz, were spaced logarithmically, and had a bandwidth of approximately 1/12 octave ($Q_{3\text{dB}} = 12$). The output of the filters is proportional to the amplitude (square root of power) of the stimulus spectrogram. To improve the signal-to-noise of STRF estimates, the output of the filter bank was smoothed across frequency and downsampled to 24 channels.

In some cases, STRF performance can be improved by applying a compressive nonlinearity to the stimulus spectrogram, to simulate processing in the auditory nerve (Gill et al. 2006). When we applied a logarithm to the stimulus spectrogram prior to STRF estimation, we observed a small although insignificant decrease in prediction accuracy for both boosted and NRC STRFs. This transformation had no effect on the significance of findings in this study. Thus the results reported here were computed without a compressive nonlinearity. A more biologically plausible transformation that includes threshold and saturation in the auditory nerve is likely to improve performance (Gill et al. 2006).

Spike rates were computed by averaging responses over the five repeated stimulus presentations. Both the stimulus spectrogram and spike rates were finally binned at 10 ms resolution.

This data preprocessing requires the selection of two additional hyperparameters, the spectral and temporal sampling density. The respective values of 1/4 octave and 10 ms were chosen to match the spectral resolution of critical bands (Zwicker 1961) and temporal resolution of neurons in A1 (Schnupp et al. 2006). Increased binning resolution would change the number of fit parameters and could, in theory, affect the performance of the two estimation algorithms. Since typical NRC estimates are substantially smoothed, even at the resolution used in this study (Figure 5), the performance of NRC is likely not to be affected by increased resolution. Changing spectro-temporal resolution might improve the performance of STRFs estimated by boosting (see ‘Discussion’). However, the resolution used in this study is sufficient for demonstrating the improved performance of boosting over NRC in these experimental conditions.

Validation procedure. A cross-validation procedure was used to make unbiased measurements of the accuracy of the different STRF models. From the entire speech data set, 95% of the data was used to estimate the STRF (estimation data set). The STRF was then used to predict the neuronal responses in the remaining 5% (validation data set), using the same 10 ms binning. This procedure was repeated 20 times, excluding a different validation segment on each repeat. Each STRF was then used to predict the responses in its corresponding validation data set. These predictions were concatenated to produce a single prediction of the neuron's response. Prediction accuracy was determined by measuring the correlation coefficient (Pearson's r) between the predicted and observed response. This procedure avoided any danger of overfitting or of bias from differences in model parameter and hyperparameter counts (David and Gallant 2005).

Studies using natural stimuli have argued that A1 encodes natural sounds with 10 ms resolution (Schnupp et al. 2006), but in some conditions A1 neurons can respond with temporal resolution on the order of 4 ms (Furukawa and Middlebrooks 2002). Measurements of prediction accuracy by correlation ignore signals at resolutions finer than the temporal window (Theunissen et al. 2001); thus the correlation values reported in this study should not be interpreted as strict lower bounds on the portion of responses predicted by STRFs.

Each STRF estimated from the speech data was also used to predict responses to the TORC stimuli. Reported prediction correlations for TORCs are averaged across the 20 STRFs estimated for cross validation.

Projection of boosted STRFs into the NRC subspace

Boosting and NRC differ in the priors they place on estimated STRFs. Boosting assumes a sparse STRF while NRC assumes that STRF power will exist only along stimulus dimensions that have high power.

Given the prior used in NRC, STRFs estimated by NRC span only part of the stimulus space. For accurate comparison of STRFs estimated using the different procedures, it is necessary to project boosted STRFs into the same subspace as the NRC STRFs (David et al. 2004; Woolley et al. 2005). For each neuron, the boosted STRF, h_b , was projected into the NRC subspace by introducing the pseudoinverse projection (David et al. 2004),

$$h_{b/\text{NRC}} = C_{\text{approx}}^{-1} C_{ss} h_b \quad (14)$$

Tuning properties derived from STRFs

To compare STRFs estimated using the different stimulus algorithms, we measured a similarity index between them by computing their normalized pointwise correlation. An index value of 1 indicated an exact match of excitatory and inhibitory subfields, while a value of 0 indicated no correlation.

We also measured four tuning properties commonly used to describe auditory neurons (Kowalski et al. 1996; Klein et al. 2000): best frequency, peak latency, spectral scale, and preferred modulation rate. Best frequency and peak latency were measured by computing the center of mass of the STRF after setting all negative

parameters to zero. Spectral scale, measured in cycles per octave (cyc/oct), is inversely proportional to spectral bandwidth; scale describes the spectral structure of a stimulus that most strongly drives the neuron. Preferred modulation rate, measured in cycles per second (Hz), is a complementary property to scale that describes the temporal modulation rate of a stimulus that best drives the neuron. Spectral scale and preferred rate were measured by computing the modulation transfer function of the STRF (i.e., the absolute value of its two-dimensional Fourier transform, (see Klein et al. 2000 and Woolley et al. 2005), averaging the first quadrant and transpose of the second quadrant, and computing the center of mass of the average.

Results

Spectro-temporal receptive fields of A1 neurons during stimulation by continuous speech

In order to study how human speech is represented in primary auditory cortex (A1), we recorded the responses of 164 isolated A1 neurons to continuous speech stimuli (Figure 1A). The stimuli were taken from a standard speech processing library and were sampled over an assortment of speakers, balanced between male and female (Garfalo 1988). We observed reliable responses to repeated stimulus presentations, often with brief bursts associated with the onset of syllables (Figure 1B).

We characterized the functional relationship between the stimulus and neuronal response by estimating the STRF for each neuron from its responses to the speech stimulus. The STRF is a linear mapping from the stimulus spectrogram to the neuron's instantaneous firing rate response (Kowalski et al. 1996; Klein et al. 2000; Theunissen et al. 2001). Often, STRFs are estimated using a linear spectrogram representation of the stimulus. In order to use a more biologically plausible model, we used a slightly different spectrogram representation that simulates the output of the auditory nerve with a bank of logarithmically spaced gammatone filters (Yang et al. 1992).

In the case of white noise stimuli (i.e., stimuli that are uncorrelated in their spectrogram representation), STRFs can be estimated simply by spike-triggered averaging (Kowalski et al. 1996; Klein et al. 2000; Theunissen et al. 2001). However, natural stimuli contain strong correlations over time and between spectral channels that require more complex algorithms to achieve unbiased STRF estimates (Theunissen et al. 2001; Singh and Theunissen 2003).

We compared two algorithms for finding the minimum mean-squared error estimate of the STRF that differ only in their prior assumptions about model parameters. The first, *boosting*, iteratively increments single STRF parameters by a small amount until further increments fail to improve response predictions (Friedman et al. 2000; Willmore et al. 2005; Zhang and Yu 2005; Willmore et al. 2006). Boosting tends to minimize the number of nonzero parameters, effectively imposing a sparse prior on the STRF (Sahani and Linden 2003). The second algorithm, *NRC*, is a variant of classical linear regression that estimates the STRF only in the stimulus subspace with high signal-to-noise (Theunissen et al. 2001). Speech and other natural stimuli tend to have $1/f^2$ power spectra (Singh and Theunissen 2003), and spectro-temporal features at high frequencies tend to be

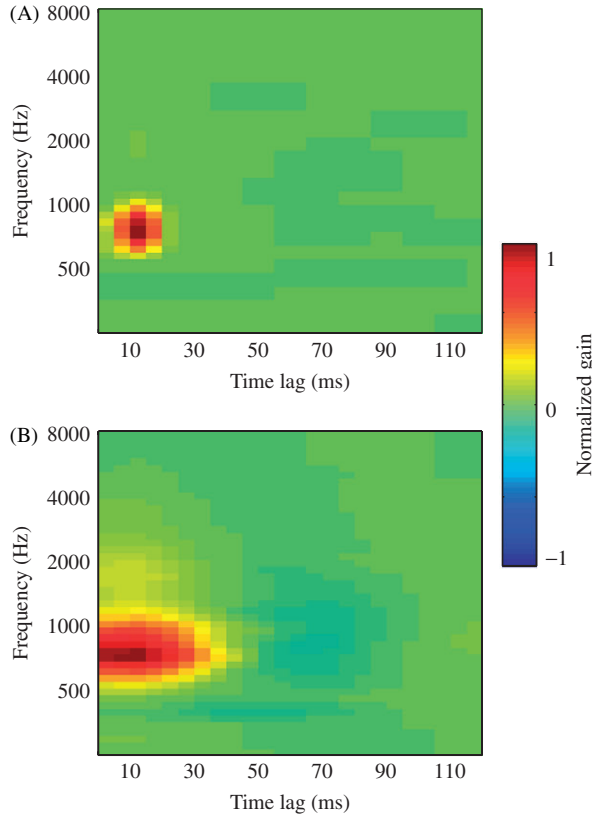


Figure 2. (A) Example of a STRF estimated from the speech stimulus by boosting. Red regions indicate frequencies and time lags for which high stimulus power is correlated with an increase in neuronal spiking. Blue regions indicate stimulus components correlated with a decrease in spiking. Boosting estimates the STRF by an iterative algorithm. During each iteration, the single parameter that best improves prediction accuracy is incremented or decremented by a small amount until further increments add noise. This procedure biases STRFs to be sparse, minimizing the number of nonzero parameters. (B) STRF estimated for the same neuron using NRC. NRC biases parameters associated with low-power dimensions of the stimulus to be zero. For natural stimuli such as speech, this introduces a smooth prior. Thus the NRC STRF has broader spectral and temporal tuning than the boosted STRF in (A). (STRFs have been upsampled by a factor of two and smoothed to facilitate visualization.)

excluded from the STRF estimated using NRC. Projection into this subspace effectively imposes a smooth prior on the STRF (Sahani and Linden 2003; Willmore and Smyth 2003).

We found that the STRFs estimated for A1 neurons from the speech data had bandpass spectro-temporal tuning, typical of measurements with other stimuli (Kowalski et al. 1996). Figure 2A shows an example of an STRF estimated by boosting. This neuron is excited by a narrow range of frequencies (best frequency 700 Hz, spectral scale 0.49 cycles per octave) with a peak latency of 13 ms and preferred modulation rate of 23.4 Hz. In order to test the accuracy of the STRF, we used it to predict the response to a validation stimulus that was also played to the

neuron but was not used for STRF estimation (David and Gallant 2005). The STRF estimated using speech predicts the validation stimulus quite accurately, with a correlation coefficient of $r=0.85$ (10 ms time bins).

The STRF estimated by NRC from the same speech data shows substantial differences from the STRF estimated by boosting (Figure 2B). Its best frequency (725 Hz) and peak latency (17 ms) are similar, but its spectral scale (0.40 cyc/oct) and preferred rate (14.3 Hz) are much lower than the boosted STRF. The lower scale and rate are reflected in the wide spectral and temporal spread of the excitatory region of the STRF in Figure 2B. Surprisingly, despite the differences in tuning, the NRC STRF predicts with nearly the same accuracy ($r=0.83$) as the STRF estimated by boosting. These predictions are not significantly different (randomized paired t -test, $p > 0.2$).

Although STRFs estimated by boosting and NRC tend to have different spectral scale and rate tuning, they generally share basic features. Figure 3 compares STRFs estimated using the two algorithms for a different neuron. The boosted STRF (Figure 3A) has an excitatory domain centered at 1160 Hz and two smaller

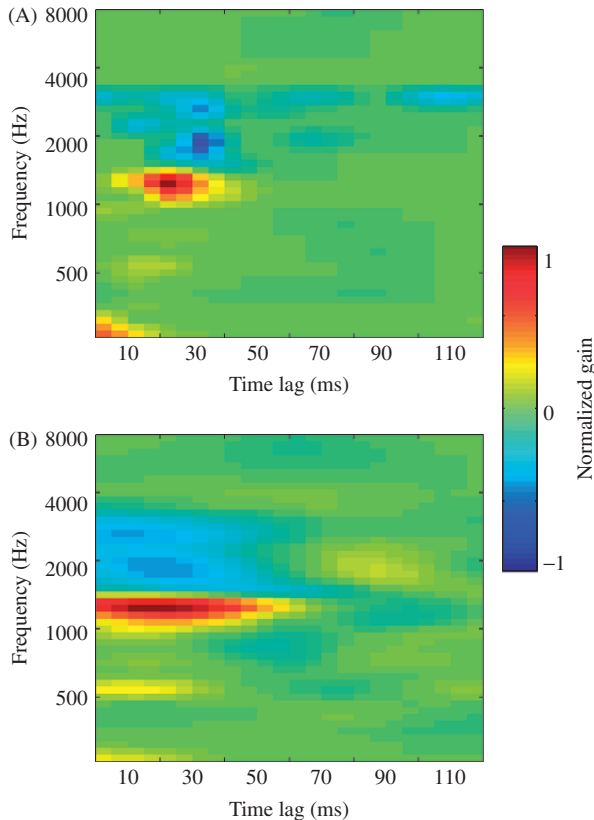


Figure 3. (A) Example STRF of a neuron with more complex tuning estimated by boosting. This STRF has an excitatory region near 1160 Hz and two inhibitory regions centered at higher frequencies. (B) The STRF estimated by NRC for the same neuron shows a similar pattern of spectral tuning as in (A), but is smoother in time.

inhibitory domains at higher frequencies, 1800 and 2500 Hz. The NRC STRF (Figure 3B) contains the same pattern of one excitatory domain at a lower frequency than two inhibitory domains. In this example, the preferred scale of the STRFs estimated by boosting (0.81 cyc/oct) and NRC (0.80 cyc/oct) are similar. However, the STRFs have different preferred rates. As in the previous example, the boosted STRF has a higher preferred rate (19.0 Hz) than the NRC STRF (11.4 Hz).

When we compared the ability of these STRFs to predict responses in a validation data set, we found, as in the previous example, that the boosted STRF ($r=0.44$) performed slightly, but not significantly better than the NRC STRF ($r=0.42$, $p>0.2$).

Comparison of prediction accuracy for STRFs estimated by boosting and NRC

We compared the ability of STRFs estimated by boosting and NRC to predict responses to speech across the entire set of 164 A1 neurons in our sample (Figure 4A). The average prediction correlation for STRFs estimated by boosting is $r=0.27$, slightly but significantly greater than the average for STRFs estimated by NRC, $r=0.25$ ($p<0.001$, randomized paired t -test, Figure 4B). Within the sample of neurons, STRFs estimated by boosting can predict validation responses with greater than random accuracy for 87% of neurons ($n=143/164$, $p<0.05$,

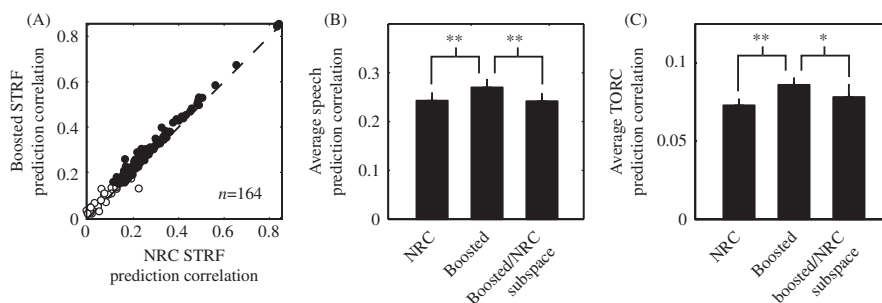


Figure 4. (A) Boosted and NRC STRFs were evaluated by their ability to predict responses to a validation speech data set that was not used for STRF estimation. Each point plots the correlation coefficient between the predicted and observed response for the NRC STRF (horizontal axis) and boosted STRF (vertical axis) for a single neuron. Black dots indicate neurons (133/164) for which both STRFs predict with greater than random accuracy (jackknifed t -test, $p<0.05$). (B) The average prediction correlation for boosted STRFs (mean 0.27) is significantly greater than the average for NRC STRFs (mean 0.24, $p<0.001$, randomized paired t -test). After the boosted STRFs are projected into the stimulus subspace used for NRC, however, prediction accuracy (mean 0.24) is significantly reduced from the original boosted STRFs ($p<0.001$, randomized paired t -test) and the difference from NRC STRFs disappears. (C) Average prediction correlation for boosted and NRC STRFs predicting responses to TORCs. The average prediction correlation for boosted STRFs (mean 0.084) is significantly greater than for NRC STRFs (mean 0.072, $p<0.001$, randomized paired t -test). As in the case of speech predictions, the difference in prediction accuracy disappears after the boosted STRFs are projected into the NRC stimulus subspace. (* $p<0.05$; ** $p<0.001$).

jackknifed t -test). Nearly the same number of STRFs estimated by NRC predict validation data with greater than random accuracy (82%, $n = 134/164$).

To evaluate how well STRFs estimated using speech generalize to other stimulus domains, we also compared how well they predicted responses to TORCs (Klein et al. 2000) played to the same set of neurons (Figure 4C). (TORCs are a convenient stimulus for estimating spectro-temporal response properties, but this study is restricted only to evaluating how well STRFs estimated using speech predict responses to TORCs.) As with the speech predictions, STRFs estimated by boosting tend to predict TORC responses (mean 0.084) better than STRFs estimated by NRC (mean 0.072, $p < 0.001$, randomized paired t -test). The absolute prediction accuracy for both sets of STRFs is substantially lower than their speech predictions. This decrease in performance suggests that neither STRF generalizes completely to other stimulus classes. Because of nonlinear response properties, STRFs estimated using one stimulus class tend to predict responses to other stimulus classes with worse accuracy (Theunissen et al. 2001; David et al. 2004). However, the superior performance of boosted STRFs suggests that they provide a more general characterization of spectro-temporal tuning across different stimulus conditions.

Effects of estimation algorithm priors on STRFs

The similar performance of the two sets of STRFs, despite their substantial differences in tuning, may seem paradoxical. It is important to consider, however, that the two estimation algorithms assume very different priors about the STRFs. When signal-to-noise levels are high, priors have relatively little effect, and the stimulus response data dominates the estimates. However, when signal-to-noise is low, priors provide an informed guess for how the STRF might appear, based on existing knowledge of the system under study.

Boosting employs a sparse prior. On each iteration, a single STRF parameter is incremented by a small amount until further increments add noise. The optimal number of iterations varies across neurons. For this study, the number of iterations to complete for each neuron was determined by cross-validation (see ‘Methods’). In the presence of noise, only a few coefficients will be nonzero; thus the STRF will be biased to be sparse. This process is illustrated for an example neuron in Figure 5A. After a few iterations, only a few parameters are non-zero. With successive iterations, more parameters are filled in and predictions improve. After the optimal number of iterations (circled in gray), additional increments add noise and predictions worsen.

NRC uses a different prior, fitting the STRF only in the subspace of the spectrum that contains most of the variance in the speech stimulus. Parameters that describe responses outside this subspace are fixed at zero. A neuron may be tuned in the excluded space, but tuning in that space has relatively little impact on responses to speech, as speech stimuli, by definition, contain little power in that space. The size of the subspace preserved by NRC is specified by a tolerance value, which corresponds to the fraction of stimulus variance to preserve (Theunissen et al. 2001). The optimal tolerance value depends on the tuning and responsiveness of individual neurons. In this study, tolerance values ranged from 90%

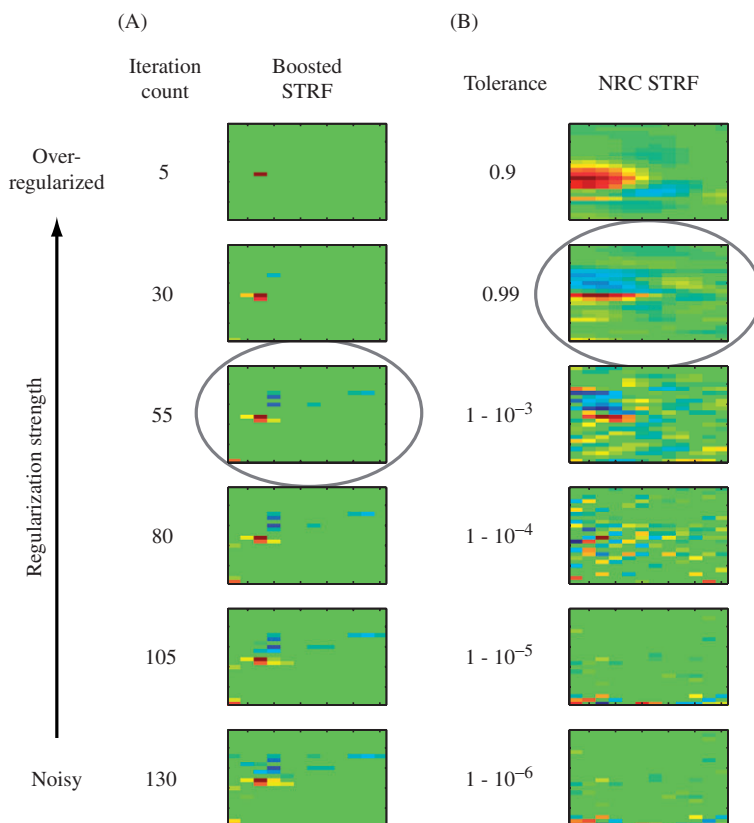


Figure 5. (A) Illustration of the sparse prior for boosting. The iterative boosting algorithm must be stopped early to avoid overfitting to noise. Panels from top to bottom show STRFs estimated by boosting after increasing numbers of iterations (STRFs have been plotted in their raw unsmoothed state). After a few iterations, only a single parameter is nonzero. Further iterations allow more nonzero parameters but also introduce noise. The optimal number of iterations (STRF circled in gray) is determined by cross validation. (B) Illustration of the smooth prior for NRC. NRC avoids overfitting to noise by projecting the fit stimulus into the subspace that contains most of its variance. The fraction of stimulus variance preserved after the projection is specified by a tolerance value. STRFs estimated using increasingly large tolerance values are shown in the panels from top to bottom. For very low tolerances, the STRF is extremely smooth, and the smaller inhibitory regions cannot be identified. Higher tolerance values reveal more detail, but at excessively high values the STRF is dominated by noise. The optimal tolerance value (STRF circled in gray) is chosen by cross validation.

(most regularized) to 99.9999% (least regularized), and the value for each neuron was chosen by cross-validation (see ‘Methods’).

Speech, like most other natural sounds, contains relatively broadband components and changes relatively slowly in time. Thus speech contains relatively little power at high spectral scales (i.e., modulation of stimulus power across frequency channels) and little power at high rates (i.e., modulation of stimulus power over time). The stimulus domain typically excluded by NRC contains these high rates and high scales, and the effect of a low tolerance value is to smooth spectro-temporal

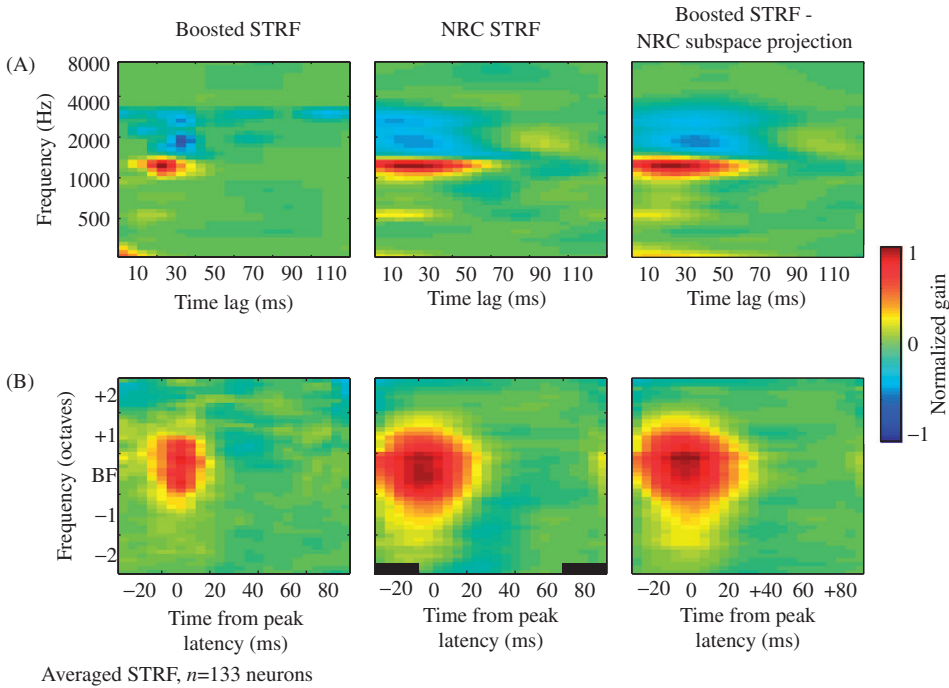


Figure 6. (A) The boosted STRF estimated for one neuron (left) shares basic spectral tuning with the NRC STRF estimated for the same neuron (center) but is more narrowly tuned in time. When the boosted STRF is projected into the subspace used by NRC (right), the tuning is nearly identical to the NRC STRF. (B) The average boosted (left) and NRC (center) STRF was calculated after shifting each STRF to have the same best frequency and peak latency and normalizing each to the same gain. The average boosted STRF has a higher spectral scale (i.e., narrower bandwidth) and a higher modulation rate (i.e., narrower latency range). When the average is computed for boosted STRFs after projecting each into the NRC subspace used for that neuron (right), it is nearly identical to the average NRC STRF.

tuning. This is illustrated for an example neuron in Figure 5B. At very low tolerance values, the STRF is highly smoothed. Reducing the tolerance tightens up tuning, but for very high tolerance values, noise begins to dominate the STRF.

Because of their different priors, the two algorithms can arrive at STRF estimates that appear very different (Figure 6A). In order to understand the source of these differences, it is useful to study the specific effects of the different priors in more detail. The relatively broadband STRF produced by NRC reflects the fact that the NRC algorithm has assumed that low-power stimulus dimensions have no effect on responses. In practice, the NRC STRF is estimated only in the high-power subspace of the stimulus, while the boosted STRF is estimated in the complete stimulus space. For comparison, the boosted STRF can be projected into the same subspace as the NRC STRF (Equation 14). When the boosted STRF is projected into this subspace, it appears much more similar to the NRC STRF (Figure 6A, right).

We can make the same comparison across the entire set of neurons in our sample. Figure 6B shows the average STRFs estimated using each algorithm for the 133/164 neurons whose STRFs predicted with greater than random accuracy (Figure 4A).

Before averaging, each STRF was shifted to have the same best frequency and peak latency and normalized to have the same overall gain. The average boosted STRF (left) clearly has narrower spectral and temporal tuning than the average NRC STRF (middle). However, when the boosted STRFs are projected into the NRC subspace, the average STRF (right) appears very similar to the average NRC STRF.

To determine the effect of projecting boosted STRFs into the NRC subspace on model accuracy, we evaluated the ability of the projected STRFs to predict the same validation data used to test the boosted- and NRC STRFs. The boosted STRFs projected into the NRC subspace predict with a mean correlation of $r=0.24$, nearly the same accuracy as NRC STRFs and significantly lower than the mean correlation for the original boosted STRFs (randomized paired t -test, $p<0.001$, Figure 4B). Thus the small enhancement in predictive power of the boosted STRFs can be attributed tuning in the subspace that is excluded by NRC.

Comparison of tuning properties of STRFs estimated by boosting and NRC

Our comparison of predictive power indicates that the superior performance of boosted STRFs lies in spectro-temporal tuning properties in the stimulus subspace excluded by NRC. To test for these differences directly, we measured a similarity index (the normalized correlation) between STRFs estimated for the same neuron using each of the methods. A similarity index of 1 indicates a perfect match, and an index of 0 indicates no correlation between STRFs. We restricted this analysis only to the 133/164 neurons for which boosted and NRC STRFs predicted with greater than random accuracy.

When measured directly between STRFs estimated using boosting and NRC, the similarity index is relatively low (mean 0.44, Figure 7A). Across the set of neurons, 111/133 (83%) have similarity indices significantly less than 1 ($p<0.05$, jackknifed t -test). However, after projecting the boosted STRFs into the NRC subspace, similarity greatly increases (mean 0.78, Figure 7B), and only 41/133 (31%) of neurons have index values significantly less than 1 ($p<0.05$, jackknifed t -test). Thus the differences in tuning observed between boosted and NRC STRFs can be attributed largely to the stimulus subspace used by NRC. The differences that persist are likely to be small effects of the sparse prior used by boosting that do not significantly affect prediction accuracy.

In order to identify specific differences in spectro-temporal tuning that give rise to the improved performance of the boosted STRFs, we compared the tuning properties of STRFs estimated using the two different algorithms. For each STRF we measured the best frequency, peak latency of excitation, spectral scale (i.e., the inverse of spectral bandwidth), and preferred temporal modulation rate.

We first compared the tuning properties of boosted STRFs directly to those of NRC STRFs. We then projected the STRF of each boosted STRF into the subspace used by NRC for that neuron and compared the tuning of the projected STRFs to NRC STRFs. The scatter plots in Figures 8A and B, respectively, illustrate the comparison of spectral scale before and after projection.

We observed only small effects of the estimation algorithm on best frequency and peak latency. The mean best frequency for boosted STRFs (2022 Hz) is slightly but not significantly greater than the mean for NRC STRFs (1986 Hz, randomized paired t -test, Figure 8C). After projecting the boosted STRFs into the NRC

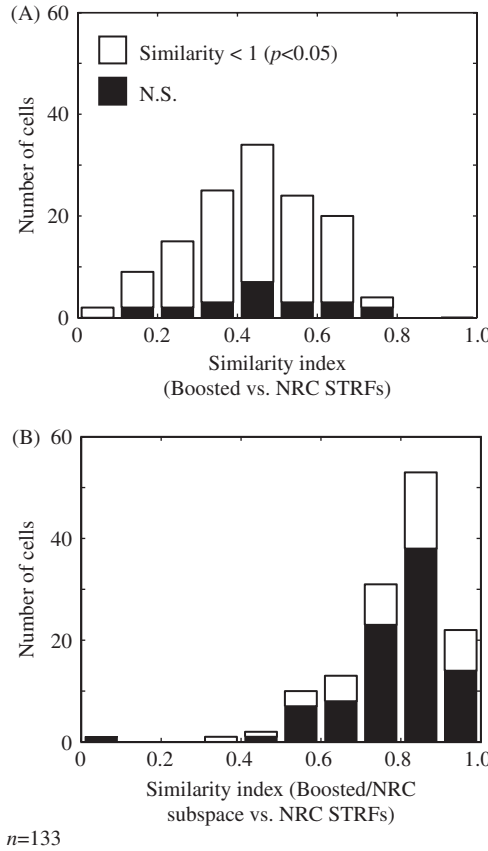


Figure 7. (A) Histogram of similarity indices computed between boosted and NRC STRFs. An index of 1 indicates an exact match between STRFs and a value of 0 indicates no correlation. The mean similarity index is 0.44. White bars indicate neurons for which similarity is significantly less than 1 ($p < 0.05$, jackknifed t -test). (B) Histogram of similarity indices computed between boosted and NRC STRFs after projecting the boosted STRFs into the stimulus subspace used by NRC. After projection, similarity increases significantly to a mean of 0.78 across the set of neurons ($p < 0.001$, randomized paired t -test).

subspace (mean 1947 Hz), the difference becomes even smaller. Mean peak latency is also not significantly different between boosted STRFs (34.3 ms) and NRC STRFs (39.0 ms, Figure 8D). The mean peak latency of boosted STRFs projected into the NRC subspace (34.5 ms) is unchanged from before the projection.

The differences in spectral scale and preferred rate determined by the estimation algorithms are more dramatic. The mean spectral scale of boosted STRFs (0.90 cyc/oct) is significantly greater than the mean for NRC STRFs (0.63 cyc/oct, $p < 0.001$, randomized paired t -test, Figure 8E). After projecting the boosted STRFs into the NRC subspace (mean 0.60 cyc/oct), the difference becomes much smaller and is no longer significant ($p > 0.2$). The mean preferred rate of boosted STRFs (22.4 Hz) is nearly twice the mean for NRC STRFs (13.4 Hz, $p < 0.001$, randomized paired t -test, Figure 8F). After projecting the

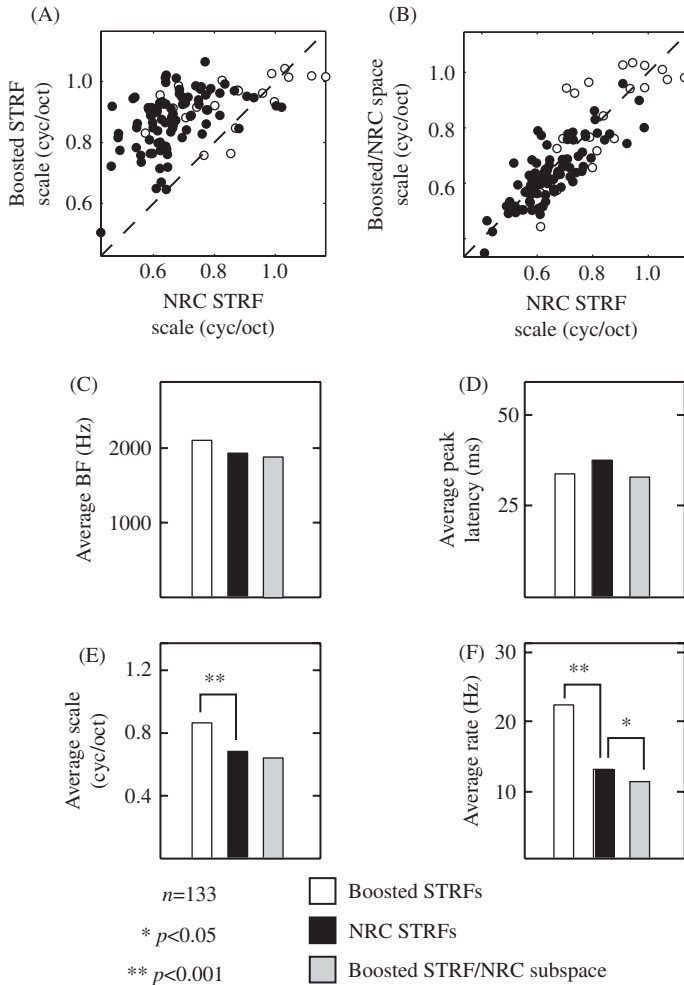


Figure 8. (A) Spectral scale measured from each NRC STRF (horizontal axis) is plotted against the spectral scale measured from the boosted STRF for the same neuron (vertical axis). Neurons for which both STRFs predict with greater than random accuracy ($p < 0.05$, jackknifed t -test) are plotted in black. The majority of points lie above the line of unity slope, indicating that scale is generally higher in boosted STRFs. (B) Comparison of spectral scale between NRC and boosted STRFs after projection into the subspace used to estimate the NRC STRFs. The majority of points lie near the line of unity slope, indicating that scale is similar in the two sets of STRFs. (C) Comparison of average best frequency for boosted STRFs (white), NRC STRFs (black) and boosted STRFs after projection into the NRC subspace (gray). There is no significant difference between the three groups of STRFs. (D) Comparison of average peak latency, plotted as in C. (E) Comparison of average spectral scale. Boosted STRFs have significantly greater scale than NRC STRFs ($p < 0.001$, randomized paired t -test). This difference disappears after the boosted STRFs are projected into the NRC subspace. (F) Comparison of average preferred rate. Boosted STRFs are tuned to significantly greater rates than NRC STRFs ($p < 0.001$, randomized paired t -test). This difference disappears after the boosted STRFs are projected into the NRC subspace, and the average rate is actually slightly lower than the NRC STRFs ($p < 0.05$).

boosted STRFs into the NRC subspace, the mean rate (11.9 Hz) is actually slightly lower than for NRC STRFs ($p < 0.05$). Thus the enhanced performance of boosted STRFs over NRC STRFs reflects differences in estimates of spectral and temporal bandwidth.

Discussion

The priors implicit to fitting algorithms can have substantial effects on STRFs estimated for auditory neurons with natural stimuli. We compared boosting (Willmore et al. 2005, 2006) and NRC (Theunissen et al. 2001; David et al. 2004), two STRF estimation algorithms that differ only in their priors, using data recorded from neurons in primary auditory cortex (A1) during the presentation of continuous speech stimuli. We found that boosting produces STRFs with slightly greater prediction accuracy than NRC. Despite their similar predictive power, STRFs estimated by boosting are tuned to higher spectral scales and higher modulation rates than STRFs estimated by NRC for the same neuron.

The differences between boosted and NRC STRFs can be linked directly to the priors employed by their respective estimation algorithms. STRFs estimated by NRC constrain parameters to be zero outside a high-variance subspace of the stimulus (Theunissen et al. 2001; David et al. 2004). When boosted STRFs are projected into that subspace, setting parameters outside the subspace to zero, their tuning properties no longer differ from NRC STRFs and their slight advantage in predictive power disappears.

The enhanced spectral and temporal precision of STRFs estimated by boosting enables them to predict responses to speech and non-speech stimuli better than STRFs estimated by NRC. These findings suggest that, in practice, boosting provides a more general description of spectro-temporal tuning properties in A1.

An additional advantage of boosting over NRC is that it avoids the computationally intensive inversion of the large stimulus autocorrelation matrix. Several nonlinear models have been developed as alternatives to the linear STRF (Prenger et al. 2004; Sharpee et al. 2004; Touryan et al. 2005). Boosting may scale well to those more complex nonlinear model regimes that require fitting more parameters. However, this proposition remains speculative, as interactions between model class, cost function, and priors are complex and difficult to predict (Wu et al. 2006).

Optimality of STRF estimation algorithms

The optimality of an estimation algorithm in one regime does not guarantee optimality in all regimes. Boosted STRFs are subject to their own prior, namely a tendency toward sparseness (Friedman et al. 2000; Zhang and Yu 2005). In the presence of noise, boosted STRFs are biased towards narrow tuning bandwidth. If data are so noisy that only one iteration of the boosting algorithm improves prediction accuracy, then only one parameter will have a nonzero value in the final estimate. This sparse estimate will bias the STRF to have tuning as narrow as the spectral and temporal binning of the data, regardless of the actual tuning of the neuron.

The magnitude of the bias imposed by the prior, for either boosting or NRC, depends on statistical properties specific to the stimulus used for characterization, the responsiveness of the neurons to that stimulus, and the amount of data available. Neither boosting nor NRC is guaranteed to provide an optimal general description of spectro-temporal tuning in all cases.

Several other algorithms have been developed for STRF estimation in the visual and auditory systems, using variants of the priors used by boosting and NRC (Ringach et al. 2002; Sahani and Linden 2003; Willmore and Smyth 2003; Machens et al. 2004; Prenger et al. 2004; Sharpee et al. 2004; Rapela et al. 2006). The method of automatic relevancy determination (ARD) uses a Bayesian approach to impose a sparse prior on STRF estimates (Sahani and Linden 2003; Prenger et al. 2004). Because boosting has not yet been cast in the Bayesian framework, a direct analytical comparison with ARD is difficult. Both methods penalize nonzero parameters that do not provide useful information about the mapping from stimulus to response; thus they are likely to produce similar STRF estimates. Both also suffer from a bias towards extremely sparse STRFs in noisy conditions.

A hybrid boosting algorithm that includes both sparse and smooth priors promises to provide a more general characterization of tuning, as has been demonstrated for ARD (Sahani and Linden 2003). The spectral and temporal sampling resolution used to define the STRF parameter space in this study are effectively hyperparameters that constrain the smoothness of boosted STRFs. Parametrically varying bin size or smoothing the stimulus spectrogram before boosting could improve the performance of STRFs. The optimal smoothing is likely to depend on the details of the stimulus and system under study (Willmore and Smyth 2003), and more complex system-specific transformations may provide even greater improvements (Willmore et al. 2005, 2006).

The direct comparison of performance used in this study is a useful tool for identifying important practical differences between STRF estimation algorithms. Similar comparisons of boosting and NRC with other algorithms will provide insight into the specific strengths and weaknesses of each one. As STRF estimation algorithms continue to be refined and tested in different systems and stimulus domains, boosting promises to provide a valuable complement to other approaches.

Acknowledgements

The authors wish to thank Jonathan B. Fritz for contributions to stimulus development and data acquisition. This work was supported by grants from the National Institutes of Health (R01 DC005779, S.A.S., and F32 DC008453, S.V.D.).

References

- Bizley JK, Nodal FR, Nelken I, King AJ. 2005. Functional organization of ferret auditory cortex. *Cereb Cortex* 15(10):1637–1653.
- David SV, Gallant JL. 2005. Predicting neuronal responses during natural vision. *Network* 16(2–3):239–260.

- David SV, Vinje WE, Gallant JL. 2004. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J Neurosci* 24(31):6991–7006.
- Friedman J, Hastie T, Tibshirani R. 2000. Additive logistic regression: A statistical view of boosting. *Ann Statist* 28(2):337–407.
- Furukawa S, Middlebrooks JC. 2002. Cortical representation of auditory space: Information-bearing features of spike patterns. *J Neurophysiol* 87(4):1749–1762.
- Garfalo J. 1988. Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. MD: National Institute of Standards and Technology.
- Gill P, Zhang J, Woolley SM, Fremouw T, Theunissen FE. 2006. Sound representation methods for spectro-temporal receptive field estimation. *J Comput Neurosci* 21(1):5–20.
- 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.
- Kowalski N, Depireux DA, Shamma SA. 1996. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76(5):3503–3523.
- Machens CK, Wehr MS, Zador AM. 2004. Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24(5):1089–1100.
- Prenger RJ, Wu MC-K, David SV, Gallant JL. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Networks* 17(5–6):663–679.
- Rapela J, Mendel JM, Grzywacz NM. 2006. Estimating nonlinear receptive fields from natural images. *J Vis* 6(4):441–474.
- Ringach DL, Hawken MJ, Shapley R. 2002. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J Vis* 2:12–24.
- Sahani M, Linden JF. 2003. Evidence optimization techniques for estimating stimulus-response functions. In: Becker S, Thrun S, Obermayer K, editors. *Advances in neural information processing systems*. Vol. 15. Cambridge, MA: MIT Press. pp 317–324.
- Schnupp JW, Hall TM, Kokelaar RF, Ahmed B. 2006. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26(18):4785–4795.
- Sharpee T, Rust NC, Bialek W. 2004. Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Comput* 16(2):223–250.
- Singh NC, Theunissen FE. 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114(6 Pt 1):3394–3411.
- Smith EC, Lewicki MS. 2006. Efficient auditory coding. *Nature* 439(7079):978–982.
- Theunissen FE, Sen K, Doupe AJ. 2000. Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331.
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. 2001. Estimating spatial temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network Comput. Neural Syst* 12:289–316.
- Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45(5):781–791.
- Willmore B, Smyth D. 2003. Methods for first-order kernel estimation: Simple-cell receptive fields from responses to natural scenes. *Network Comput Neural Syst* 14:533–577.
- Willmore B, Prenger RJ, Gallant JL. 2005. Spatial and temporal receptive field properties of neurons in area V2. Program No. 618.17. Washington, DC: Society for Neuroscience.
- Willmore B, Prenger RJ, Gallant JL. 2006. Spatial and temporal receptive field properties of neurons in area V4. Program No. 64.4. Atlanta, GA: Society for Neuroscience.
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE. 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8(10):1371–1379.
- Wu MC-K, David SV, Gallant JL. 2006. Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 12(4):477–505.
- Yang X, Wang K, Shamma S. 1992. Auditory representations of acoustic signals. *IEEE Trans Info Theory* 38(2):824–839.
- Zhang T, Yu B. 2005. Boosting with early stopping: Convergence and consistency. *Annu Stat* 33(4):1538–1579.
- Zwicker E. 1961. Subdivision of the audible frequency range into critical bands. *J Acoust Soc Am* 33(2):248.