

## Multivariate receptive field mapping in marmoset auditory cortex

Artur Luczak<sup>a,b</sup>, Troy A. Hackett<sup>d,e</sup>,  
Yoshinao Kajikawa<sup>e</sup>, Mark Laubach<sup>a,c,\*</sup>

<sup>a</sup> *The John B. Pierce Laboratory, New Haven, CT, USA*

<sup>b</sup> *Department of Computer Science, Yale University, New Haven, CT, USA*

<sup>c</sup> *Department of Neurobiology, Yale School of Medicine, 290 Congress Ave, New Haven, CT 06519, USA*

<sup>d</sup> *Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA*

<sup>e</sup> *Psychology Department, Vanderbilt University, Nashville, TN, USA*

Received 27 October 2003; received in revised form 23 December 2003; accepted 23 December 2003

### Abstract

We describe a novel method for estimation of multivariate neuronal receptive fields that is based on least-squares (LS) regression. The method is shown to account for the relationship between the spike train of a given neuron, the activity of other neurons that are recorded simultaneously, and a variety of time-varying features of acoustic stimuli, e.g. spectral content, amplitude, and sound source direction. Vocalization-evoked neuronal responses from the marmoset auditory cortex are used to illustrate the method. Optimal predictions of single-unit activity were obtained by using the recent-time history of the target neuron and the concurrent activity of other simultaneously recorded neurons ( $R: 0.82 \pm 0.01$ ,  $\sim 67\%$  of variance). Predictions based on ensemble activity alone ( $R: 0.63 \pm 0.18$ ) were equivalent to those based on the combination of ensemble activity and spectral features of the vocal calls ( $R: 0.61 \pm 0.24$ ). This result suggests that all information derived from the spectrogram is embodied in ensemble activity and that there is a high level of redundancy in the marmoset auditory cortex. We also illustrate that the method allows for quantification of relative and shared contributions of each variable (spike train, spectral feature) to predictions of neuronal activity and describe a novel “neurolet” transform that arises from the method and that may serve as a tool for computationally efficient processing of natural sounds.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Spectrotemporal receptive field; Auditory neurons; Natural sounds

### 1. Introduction

The auditory cortex encodes spectral components of complex, biologically relevant sounds (Rauschecker et al., 1995; Recanzone et al., 1999; Wang, 2000) as well as other acoustic features such as sound source direction (Kelly et al., 2003; Middlebrooks et al., 1994, 1998), temporal coherence (Liang et al., 2002), and intensity (Schreiner et al., 1992). The response properties of neurons in the auditory cortex are commonly assessed using methods for spectrotemporal receptive field (STRF) analysis (e.g. Aertsen and Johannesma, 1981; Escabi and Schreiner, 1999; Sen et al., 2001; Shamma et al., 1995). The STRF quantifies the frequency tuning of

neurons as a function of time after the onset of the stimulus. The most commonly used method for estimating receptive fields is reverse correlation (e.g. de Boer and Kuyper, 1968; DeAngelis et al., 1995; Escabi and Schreiner, 1999; Shamma et al., 1995). This method results in a spike-triggered average of a windowed spectrogram that approximates the STRF and quantifies the correlation between neuron activity and sound frequencies as a function of time (de Ruyter van Stevenick and Bialek, 1988; Rieke et al., 1997). The reverse correlation method does not correct the STRF for correlation within a given stimulus. For this reason, the method can be used only with simple, uncorrelated stimuli, such as white noise or random tones. Only in limited instances have researchers been able to design more realistic stimuli that remain suitable for reverse correlation analysis such as evenly distributed, rapidly presented random chords (deCharms et al., 1998) or dynamic ripples (Depireux et al., 2001). To account for

\* Corresponding author. Tel.: +1-203-562-9901; fax: +1-203-624-4950.

E-mail address: mark.laubach@yale.edu (M. Laubach).

more realistic stimuli, Theunissen and Doupe (1998) proposed a method that uses a stimulus autocorrelation matrix and that requires transforming the neuronal responses into the frequency domain. This latter aspect of the algorithm may not be suitable for signals with sharp transitions (e.g. a delta function) such as neuronal spike trains. Moreover, Willmore and Smyth (2003) pointed out that the method from Theunissen and Doupe (1998) discards information about the phase of the stimulus and this is likely to be a major limitation for analyses of realistic acoustic stimuli.

Auditory neurons are well known to exhibit non-linear responses to spectrotemporal features of acoustic stimuli (e.g. Calhoun and Schreiner, 1998). Therefore, the reliance of existing methods for calculating the STRF on linear transfer functions and standard Fourier transform (as in the study by Theunissen and Doupe, 1998) has resulted in relatively poor performance in predicting neuronal responses. For example, Sahani and Linden (2003) calculated that in primary auditory cortex of rodents, the STRF can account for no more than 40% of the stimulus-related power. This limitation is also reflected in the low correlation coefficients ( $R$ ) between the actual neuronal response and the predicted response usually reported for STRF-based analyses (e.g. in Sen et al., 2001,  $R$  was in the range from 0 to 0.7). A new approach is needed that can address these issues as well as allow for a wider range of signal attributes to be examined. In the method described here, we have expanded the standard STRF approach to allow for the examination of a wide range of factors that might contribute unique information to the activity of neurons in the auditory cortex. This new method is able to quantify the relative and shared contributions of each factor to predictions of neuronal activity. By including non-spectral variables (e.g. time after stimulus onset, sound source direction), we are able to improve predictions on vocalization-evoked neuronal responses over that provided by traditional STRF analysis.

## 2. Methods

### 2.1. Electrophysiology

All procedures were approved by the IACUC at Vanderbilt University. Neural recordings were obtained in two anesthetized (ketamine hydrochloride, 10 mg/kg and xylazine, 2 mg/kg; I.M.) marmoset monkeys (*Callithrix jacchus jacchus*) using standard neurophysiological methods. Recordings were obtained using a multi-channel acquisition system (Tucker-Davis Technologies, Gainesville, FL), controlled by Brainware software developed by Jan Schnupp (Tucker-Davis Technologies). Linear electrode arrays (1 × 4 configuration; 1 mm spacing) were constructed from polyamide insulated tungsten electrodes (1.0 MΩ impedance; 1 μm tip; 0.254 mm shaft diameter) (Micro-Probe, Potomac, MD). The array was advanced by a microdrive (David Kopf, Tujunga, CA) at an angle per-

pendicular to the pial surface of cortex. Electrode depth ranged from 750 to 1000 μm, targeting responsive neurons in cortical layer III. Neuronal spike trains were recorded simultaneously from the four electrodes located in area AI of auditory cortex (Hackett et al., 2001). Data were collected from 17 separate penetrations. The recordings consisted of clearly resolved single-unit and multiple-unit clusters containing spikes from several neurons that could not be resolved using standard methods for spike sorting (i.e. thresholding and PCA). Because some of the units had very low activity, we selected the 23 units with the highest firing rates for the analyses described here. The analyses described in this manuscript were based on peri-stimulus time histograms with 10 ms bins. Additional analyses were carried out using bin sizes of 2, 5, and 15 ms bins and the results obtained were qualitatively and quantitatively similar to the results described here.

### 2.2. Acoustic stimuli

Experiments were conducted in a sound-isolating chamber (Industrial Acoustics Corp., New York, NY) located within the auditory research laboratory at Vanderbilt. Auditory stimuli were presented using Tucker-Davis Technologies System II hardware and software. Stimuli were calibrated using a 1/4 in. microphone system (ACO Pacific, Belmont CA) and SigCal software (Tucker-Davis). In this study, we analyzed responses to three exemplars of the marmoset monkey twitter call provided by Wang (Johns Hopkins University). The R.M.S. amplitude of each call was adjusted to 60 dB SPL. Each of the three calls was presented 15 times. The length of the twitters ranged from 0.8 to 1.7 s (a sample spectrogram of a twitter is shown in Fig. 1).

### 2.3. Multivariate receptive field model

The linear response function of a neuron can be defined in matrix notation as:

$$\mathbf{S} \times \mathbf{C} = \mathbf{R} \quad (1)$$

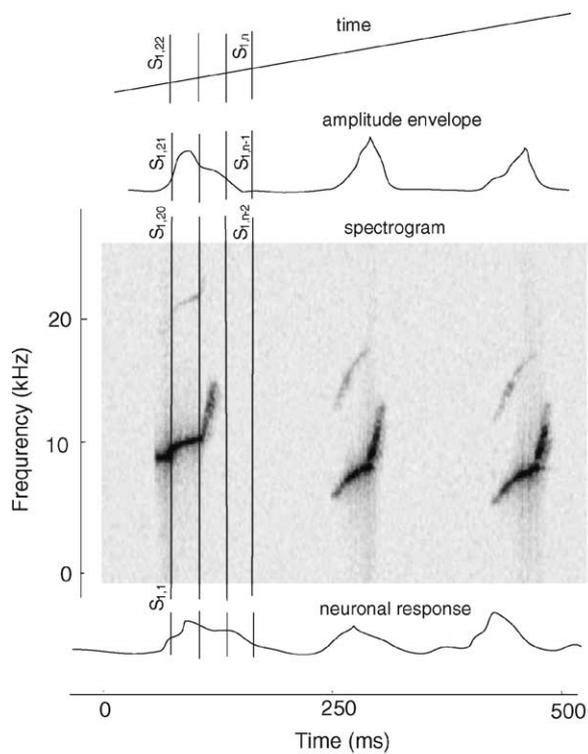
where  $\mathbf{R}$  is the response of a neuron (mean firing rate evaluated from peri-stimulus time histogram);  $\mathbf{C}$ , evaluated response function (e.g. STRF), and  $\mathbf{S}$ , matrix of stimulus features, including, e.g. the spectrogram of the sound, the amplitude of the sound envelope, the sound source direction, the derivative of the spectrogram, etc. This is schematically shown in Fig. 1, where each row in  $\mathbf{S}$  represents one time window of the stimulus. The least-squares (LS) method was applied to solve this equation. Details of the LS implementation are provided in Appendix A.

The coefficients,  $\mathbf{C}$ , evaluated from the LS model have an interesting property. If we normalize the variables in the  $\mathbf{S}$  matrix (spectrogram, amplitude, etc.) to mean zero and unit variance, then the  $\mathbf{C}$  matrix has coefficients in the range of -1 to 1 and express the relative contribution/importance of each variable. For example, if a coefficient in  $\mathbf{C}$

corresponding to a stimulus variable in  $S$  would have a value of  $-0.8$  then this variable has a large inhibitory contribution to the neuronal response. This results in a means of making direct comparisons between different types of variables that may underlie the relationship between acoustic stimuli and neuronal activity.

#### 2.4. Preprocessing

Prior to the multivariate receptive field analysis, principal component analysis (PCA) was used for dimension reduction. For the monkey calls, PCA reduced the complexity of log-transformed spectrograms (cf. Theunissen and Doupe, 1998) down to 15% of its original size while preserving 98% of signal variation (i.e. 40 principal components (PCs) represented a window of a spectrogram with 26 frequency



Multivariate Least Squares Model

$$\begin{pmatrix} S_{1,1} & \dots & S_{1,20} & S_{1,21} & S_{1,22} & \dots & S_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ S_{m,1} & \dots & S_{m,20} & S_{m,21} & S_{m,22} & \dots & S_{m,n} \end{pmatrix} * \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} = \begin{pmatrix} R_1 \\ \vdots \\ R_m \end{pmatrix}$$

Fig. 1. Illustration of the multivariate receptive field method. The first row in the first matrix contains values of stimulus  $S$  in the analyzed time window. This matrix contains the spectrogram, time and temporal envelope of the sound and can be expanded by adding other variables as a column vectors.  $R_1$ – $R_m$  denotes the neuronal response at  $m$  times. The LS model calculates the response function coefficients  $C_1$ – $C_n$  which can be interpreted as the multivariate receptive field.

bands (0–25 kHz) and a length equivalent to 100 ms (10 ms time bin)). PCA was also used to preprocess the neuronal signals. The activity of 22 units (predicted unit was excluded) was represented by nine PCs (91% of signal variance captured).

#### 2.5. Prediction of responses

To evaluate the performance of the method for STRF estimation, we used five-fold cross-validation, i.e. we calculated the STRF using 80% of our data (“training” data) and computed the predicted firing rate for the remaining 20% of the signal (“testing” data). After Theunissen et al. (2000),

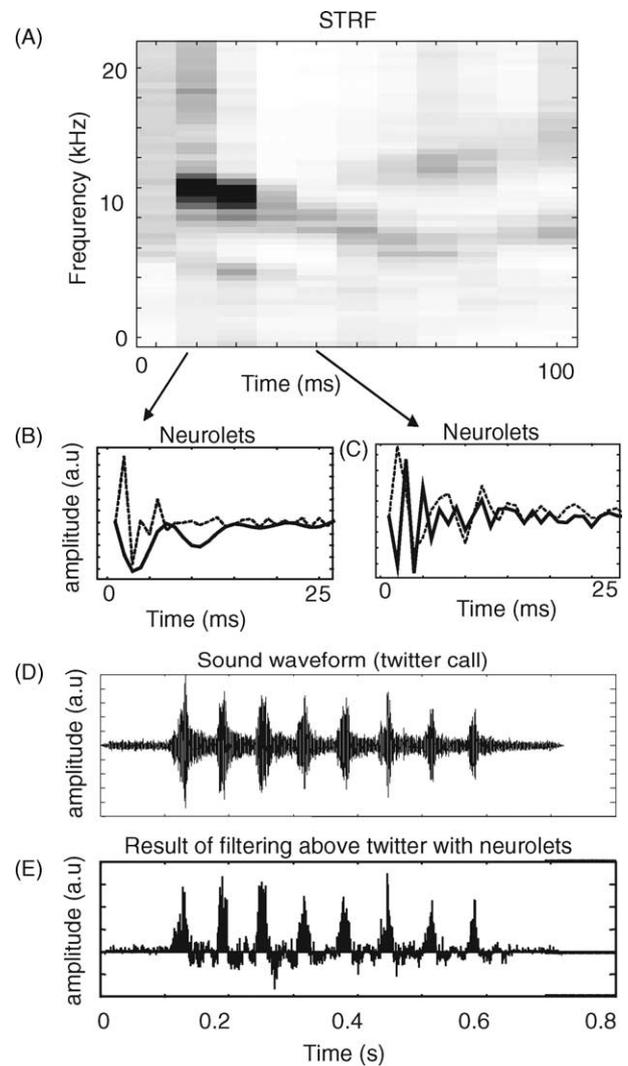


Fig. 2. Neurolets are the inverse Fourier transform of the STRF. Examples of excitatory (dashed line) and inhibitory (solid line) neurolets created from the first (B) and fourth (C) column of the STRF(A). For example, the STRF at time 10 ms (first column) indicates that this neuron, 10 ms after stimuli onset, is inhibited by low frequencies. Therefore, the corresponding inhibitory neurolet (panel B: dashed line) has strong low frequency components. D: sound waveform (twitter call). The signal in panel E corresponds to the results of filtering the twitter call using neurolets derived from the STRF above and is the predicted response of the neuron.

we used the correlation coefficient,  $R$ , to compare our predictions with neuron's actual response and converted this metric into the percent of variance explained,  $R^2$ .

### 2.6. Neurolets

A neurolet is a temporal representation of the STRF, constructed by applying the inverse Fourier transform (iFT) to the STRF. Multiplication of an acoustic signal by neurolets produced the approximate response of a neuron, the STRF of which was used to calculate the neurolets. To represent the inhibitory and excitatory parts of the STRF, we applied the iFT separately to its positive and negative parts. By so doing, we obtained two waveforms representing the two sounds that excite and inhibit, respectively, a given neuron the most (Fig. 2). We call these new representations “neurolets,” as they are similar in many ways to wavelets (Wickerhauser, 1994).

## 3. Results

### 3.1. Neuronal responses

We obtained neuronal responses from 17 locations in the auditory cortex of two marmoset monkeys. Fig. 3 illustrates examples of trial-by-trial (rasters) and average (histograms) neuronal responses from two locations in the auditory cortex. As can be seen, neuronal activity is strongly modulated by the vocal calls and each neuron is modulated in a unique manner.

### 3.2. Validation of STRF calculation

In order to test the performance of our model in a control situation, an artificial neuronal response was generated for the vocal calls. We used an “integrate and fire” model (INF) of a neuron (Salinas and Sejnowski, 2002). Original and recovered STRF are shown in Fig. 4A and B. In Fig. 4C, the spike-triggered average is given. Fig. 4D and E display examples of the raster plots of the observed and the simulated neuron's responses for a twitter. In both cases the mean firing rate was the same.

### 3.3. Multivariate receptive field analysis—single neurons

To examine whether any acoustic, physiological, etc. variable explains the non-linear components of the neuronal responses to the vocal calls, we extended our predictions of neuronal responses by including those variables in our analysis. In this work, we used: spectrograms, derivatives of spectrograms in time, envelope of sound amplitude, time course, activity of other neurons, past activity of the analyzed neuron in preceding time or different combinations of the above factors.

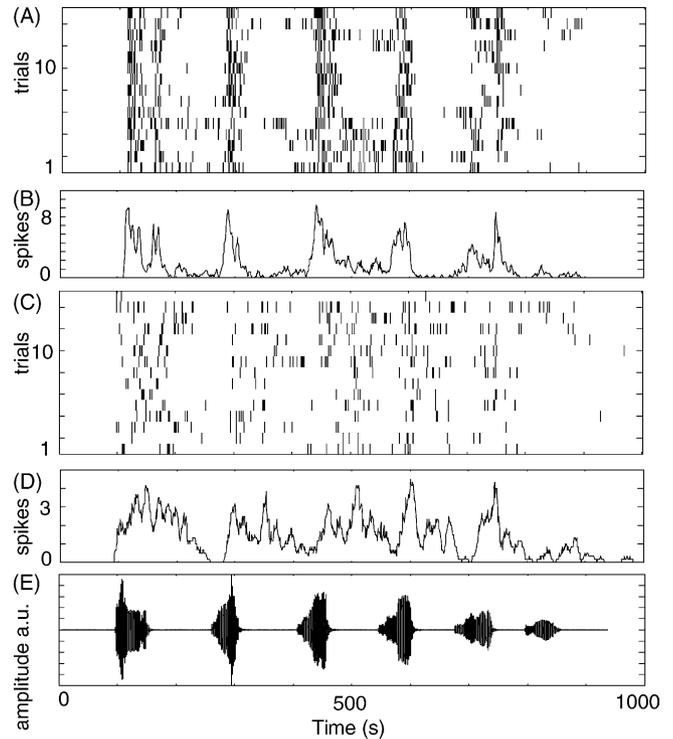


Fig. 3. Neuronal responses from the auditory cortex. Examples of spike raster plots (A, C) and peri-stimulus time histograms (B, D) from neurons located in the core area of auditory cortex. E: the time waveform of a twitter call is shown at the bottom.

To analyze the spectrograms of the vocal calls, we calculated the ‘classical’ STRF, i.e. demonstrate a neuron's response function to the frequency content of sound as in a spectrogram. The correlation coefficient ( $R$ ) was used to estimate the similarity between the neuron's predicted and actual response. The results summarized for 23 units are shown in Fig. 6 (solid line). The mean value of  $R$  was  $0.43 \pm 0.06$  S.E.M. (range:  $-0.13$  to  $0.81$ ) and so the analysis accounted for  $\sim 18.5\%$  of variance. This result is similar to the range of values reported by Theunissen et al. (2000).

Next, we examined the role of the derivatives of the spectrograms in improving predictions over the ‘classical’ STRF analysis. This feature produced a consistent decrease of  $R$  by  $0.1$ – $0.2$  (Fig. 6, dashed line). The distribution of values for  $R$  (mean:  $0.24 \pm 0.05$  S.E.M.; range:  $-0.24$  to  $0.67$ ) was larger than zero, indicating that neurons were responsive to the speed of relative changes of sound intensity (represented by the derivative of the spectrogram). Predictions based on the spectrogram derivatives thus accounted for  $\sim 5.8\%$  of variance.

To investigate if the derivative of the spectrogram provided unique information as compared to the use of the spectrogram alone, we compared predictions of the neuronal response based on the spectrogram and on the combination of the spectrogram and its derivative. There was no difference between values of  $R$  for these measures ( $R$ :  $0.43 \pm 0.27$ ; paired  $t$ -test,  $t = 0.03$ ,  $P > 0.05$ ; Fig. 7 and Table 1). There-

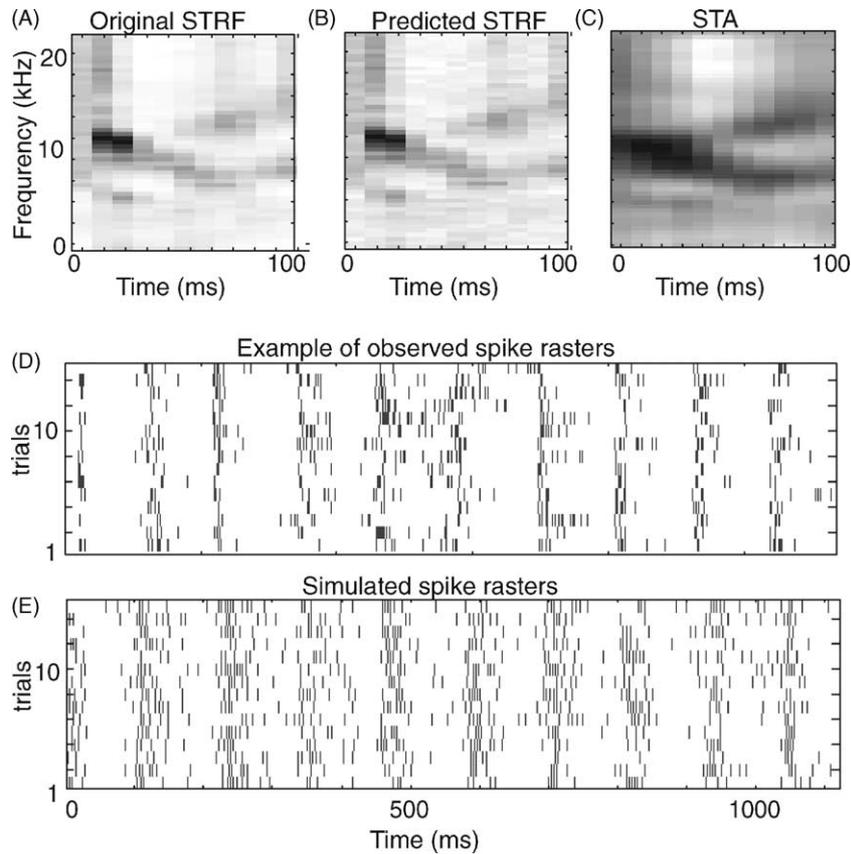


Fig. 4. Validation of the LS method: (A) the original STRF, (B) the predicted STRF, (C) the spike-triggered average (STA). Examples of raster plots: (D) a neuron's response to a twitter; (E) results of simulated INF neuron's response to the same twitter. The simulated neuron's response is less concentrated in time because it is a linear function of the spectrogram. In contrast, the actual neuron's response is not linear.

fore, information related to sound intensity changes could be obtained directly from the spectrogram.

Neuronal responses to twitters decrease with time after stimulus onset (Fig. 5). The time-dependence of the STRF was addressed by adding a vector with information about the time course of each twitter. This manipulation resulted in marginal improvements in predictions for most neurons ( $R: 0.50 \pm 0.27$ ; paired  $t$ -test,  $t = 0.88$ ,  $P > 0.05$ ; Fig. 7 and Table 1). Replacing the linear function of time course with

different non-linear curves (e.g. exponential decay) did not further improve the results (data not shown).

Finally, to address the role of the amplitude of the sound envelope, which is known to correlate with neuronal activity in the marmoset monkey (Srikantan et al., 2002), the amplitude of the sound envelope was used as a variable in the regression analysis. The amplitude envelope was calculated from the temporal representation of the sound waveform, therefore, it has a non-linear relationship to the

Table 1  
Summary of correlation coefficients calculated for predicted and actual neural responses

|         | Spec  | Deriv | Spec<br>+ deriv | Spec<br>+ ampl | Spec<br>+ time | Ensemble | Spec<br>+ ensemble | Past | Spec<br>+ past | Past<br>+ ensemble | Spec + past<br>+ ensemble |
|---------|-------|-------|-----------------|----------------|----------------|----------|--------------------|------|----------------|--------------------|---------------------------|
| Minimum | -0.13 | -0.24 | -0.17           | -0.14          | -0.14          | 0.20     | 0.05               | 0.54 | 0.53           | 0.71               | 0.60                      |
| Mean    | 0.43  | 0.24  | 0.43            | 0.43           | 0.50           | 0.63     | 0.61               | 0.73 | 0.76           | 0.82               | 0.80                      |
| Maximum | 0.81  | 0.67  | 0.82            | 0.81           | 0.81           | 0.88     | 0.87               | 0.82 | 0.85           | 0.90               | 0.89                      |
| S.D.    | 0.27  | 0.26  | 0.27            | 0.27           | 0.27           | 0.18     | 0.24               | 0.06 | 0.08           | 0.05               | 0.07                      |
| S.E.M.  | 0.06  | 0.05  | 0.06            | 0.06           | 0.06           | 0.04     | 0.05               | 0.01 | 0.02           | 0.01               | 0.013                     |

For the predictions, the following information was used: spectrogram of twitter (spec), derivative of spectrogram (deriv), spectrogram and its derivative (spec + deriv), spectrogram and amplitude of temporal envelope of sound (spec + ampl), spectrogram and time course of twitter (spec + time), activity of 22 neurons (ensemble), spectrogram and activity of 22 neurons (spec + ensemble), neuron's past activity (past), spectrogram and neuron's past activity (spec + past), neuron's past activity and activity of other neurons (past + ensemble), spectrogram and neuron's past activity and activity of other neurons (spec + past + ensemble).

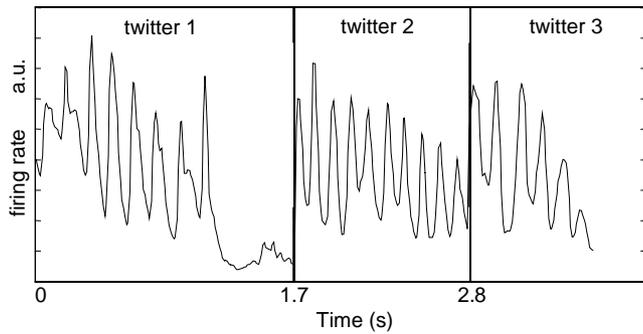


Fig. 5. The sum of the PSTHs of the 23 neurons to the three twitters demonstrates a decrease of neuronal activity with time. Vertical lines indicate the beginning of the next twitter.

log-transformed spectrogram. Predictions of neuronal responses based on the combination of this feature and the spectrogram were no better than those made with the spectrogram alone ( $R: 0.43 \pm 0.27$ ; paired  $t$ -test,  $t = 0.01$ ,  $P > 0.05$ ; Fig. 7 and Table 1).

### 3.4. Multivariate receptive field analysis—neuronal ensembles

With our method, we can estimate the response of a single neuron as a linear function of responses from other simultaneously recorded neurons. By comparing this result with the predictions made using the spectrogram alone, we examined how much information about the spectral representation of sound may be present across multiple sites in the auditory cortex. This analysis is shown in Fig. 6, where the dotted line depicts the  $R$  for predictions obtained by constructing the receptive field based on responses of the ensemble of neurons (mean:  $0.63 \pm 0.04$  S.E.M.; range: 0.2–0.88). For comparison, the solid line shows  $R$  for predictions made with the spectrogram alone. Interestingly, predictions based on the ensemble activity alone were roughly equivalent to those based on the ensembles and the spectrogram ( $R$  for ensemble:  $0.63 \pm 0.18$ ;  $R$  for ensemble and spectrogram:  $0.61 \pm 0.24$ ; see Fig. 7). This result suggests that predictive information derived from the spectral features of sound is simultaneously present in the activity of the neuronal ensemble.

### 3.5. Multivariate receptive field analysis—time history of spiking

Correlation coefficients between a neuron's current response and its response in the preceding 10 ms are shown in Fig. 7. The large value of  $R$  ( $0.73 \pm 0.01$  S.E.M.) reflects the fact that neurons' responses depend to a large extent on their past activity. By including spectrograms and past activity our predictions improved (Fig. 7 and Table 1), the value of mean  $R$  increased from 0.43 (for spectrograms alone) to 0.76 ( $t$ -test  $t(44) = 5.82$ ,  $P < 0.05$ ). Overall, we obtained the highest mean correlations by including a neuron's past ac-

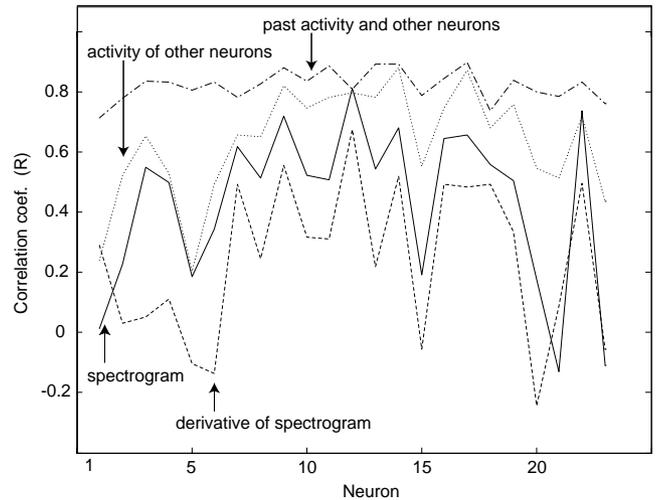


Fig. 6. Correlation coefficients calculated for predicted and actual neural responses. For the predictions, we used the following information: spectrogram (solid line), derivative of spectrogram (dashed line), activity of 22 neurons (dotted line), neuron's past activity and simultaneous activity of other neurons (dash-dot line). By using responses from other neurons we obtained significant improvement in the prediction compared to predictions from the spectrogram only (paired  $t$ -test,  $t = 2.36$ ,  $P < 0.05$ , compare solid and dotted lines). Further improvements were made by using the time-history of the target neuron and the concurrent activity of other simultaneously recorded neurons (paired  $t$ -test,  $t = 5.82$ ,  $P < 0.05$ ; dash-dot line). Replacing the spectrogram by its derivative reduced predictions (paired  $t$ -test,  $t = 2.42$ ,  $P < 0.05$ ; dashed line).

tivity and other neuron's responses ( $R: 0.82 \pm 0.01$  S.E.M.; range: 0.71–0.90; Fig. 6 (dash-dot line)) and for spectrograms and past activity and other neuron's responses (the difference between these cases was not significant; paired  $t$ -test,  $t = 1.16$ ,  $P > 0.05$ ). Adding any other additional information did not improve our results, indicating that those values of  $R$  are probably close to the upper limit of predictability for our dataset.

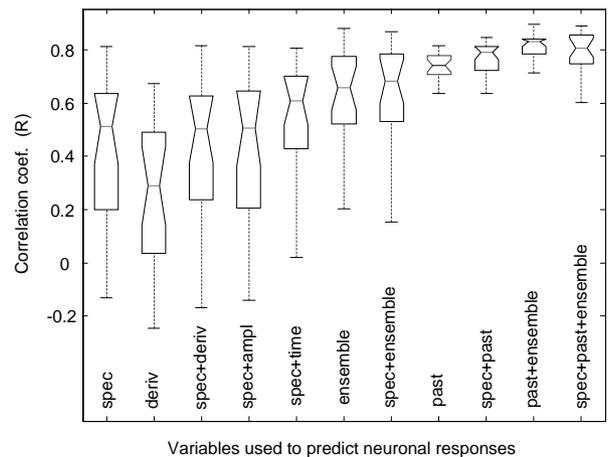


Fig. 7. Summary plot of correlation coefficients calculated for predicted and actual neural responses (the abbreviations are described in Table 1). The single box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data.

### 3.6. Neurolets

The filters derived with our method for multivariate receptive field analysis allow for predictions of neuronal responses to novel acoustic stimuli in the same way as the STRF. The difference between the ‘classical’ STRF and our filters, which we call ‘neurolets,’ is that the STRF-based predictions require conversion of the sound waveform into the time-frequency domain. Neurolets, by contrast, operate directly on the temporal representation of a sound, thus approximating the neuronal response faster and with a much smaller computational cost. We found that convolving marmoset vocalizations with neurolets acts as a detector of specific periods within the vocalizations to which the neurons are especially responsive (Fig. 2E). Consequently, neurolets can be used as a denoising algorithm suppressing those parts of a signal that do not evoke a neuronal response (Donoho, 1995; Quatieri, 2002). Neurolets derived from excitatory (positive) parts of the STRF should maximally excite the neuron. In contrast, neurolets derived from inhibitory parts of STRF represent sounds that should inhibit the neuron.

## 4. Discussion

A novel method for multivariate spectrotemporal receptive field analysis based on least-squares regression is described in this manuscript. The LS method is well understood and has successfully been applied in a variety of applications (e.g. Lawson and Hanson, 1995). The advantages of the LS method are: speed of computation, thorough understanding of the method and its limitations, simplicity of use, but the most interesting from our point of view is the possibility to analyze, in the same equation, arbitrary sets of different factors (e.g. spectrogram + time course + activity of other neurons). This is important because it gives us a tool for studying the contribution of different factors to a neuron’s response.

The application of our new method to spike train data from the marmoset auditory cortex revealed the importance of features other than time-varying spectral aspects of natural sounds such as vocal calls in modulating single neuron discharge. Traditional analyses with STRF methods that only account for the spectral content of sounds cannot reveal these relationships. We tested the relative gain/loss in predictability of neuronal responses as a function of several different acoustic (spectrogram, its derivative, amplitude envelope, time after stimulus onset) and neuronal (time-history, concurrent ensemble activity) features. This comparison uncovered a major role for the time-history of the spike train in conveying information about the current responsiveness of neurons to vocal calls. Also, our findings indicate a high level of redundancy in the auditory cortex. We could estimate the response of a single neuron as a linear function of the responses from the other units recorded in semi-random places over 4 mm of cortex.

The main limitation of the LS method proposed here is that it works with over-determined systems, i.e. the number of observations must be larger than the number of variables. This implies compromising temporal resolution or the length of the STRF. Fortunately, often this problem can be substantially overcome by dimensionality reduction of the raw data. In our case, we used principal component analysis (as described in the Section 2). By this method we can evaluate the linear contribution of the spectrogram to a neuron’s response more robustly because PCA-based dimension reduction regularizes stimulus space and reduces noise. Nevertheless this representation may not be optimal for non-stationary signals. Natural sounds seem to be very diverse but surprisingly they have similar statistics (Attias and Schreiner, 1997; Lewicki, 2002). Alternatives to PCA, such as independent component analysis (review on ICA estimation methods: Hyvärinen, 1999), may be better suited to dealing with the statistics of natural sounds and thus may better serve for preprocessing in quantitative receptive field analyses.

In addition to providing insight into neuronal responses to stimuli used in receptive field mapping experiment, the method used for data analysis in this manuscript is able to define filters based on the multivariate regression analysis. We call these filters ‘neurolets’ (Fig. 2). The difference between neurolets and the ‘classical’ STRF is that the STRF requires conversion of a sound waveform into the time-frequency domain. Neurolets, by contrast, operate directly on the temporal representation of a sound, thus approximating neuronal responses faster and with a much smaller computational cost. Such filters are a novel, brain-inspired means of processing sounds. It has been postulated that neurons code natural sounds efficiently (Lewicki, 2002). Therefore, neurolets could be useful as bases for efficient (compressed) representation of natural sounds, in a similar way as wavelets (Wickerhauser, 1994). Another interesting feature of those filters is that neurolets can be used to design the optimal stimulus for a neuron. The neurolets derived from excitatory parts of STRF are equivalent to the most preferred sound for a neuron, the STRF of which was used to calculate the neurolets.

The application of multivariate receptive field analysis to the activity of neuronal ensembles represents a significant extension of ongoing efforts aimed at understanding how complex sounds, such as natural vocalizations, are represented in the auditory cortex of primates (Wang, 2000; Wang et al., 1995). Single neurons in primary auditory cortex respond to a wide range of vocal calls and complex sounds. Yet, while natural vocalizations evoke stronger responses in primary auditory cortex compared with other complex sounds, including time-reversed vocal calls, single neurons do not function as “call detectors.” Rather, the encoding of the acoustic vocal calls appears to be accomplished by spatially distributed neuronal populations. The methods described here support those findings, and could provide valuable new insights into the encoding mechanisms that con-

tribute to the representation of complex sounds in auditory cortex.

## Acknowledgements

We thank Dr. Xiaoqin Wang for marmoset vocalizations used here and Dr. Emilio Salinas for making code for his integrate-and-fire neuron models available over the Internet. We also thank Dr. William R. D'Angelo, Dr. Susanne J. Sterbing-D'Angelo and Linda Pelechacz for helpful comments on the manuscript. This work is supported by a contract from DARPA (N66001-02-C-8023), funds from the John B. Pierce Laboratory to M.L., and by NIH/NIDCD R01 DC04318 to T.A.H.

## Appendix A

Sample MATLAB code demonstrating usage of the LS method for evaluating multivariate receptive fields is given below. In this example, the spectrogram is preprocessed with PCA. The input data for the MATLAB m file is the neural response (PSTH), the spectrogram, and the time course of stimulus. The output of the function is the multivariate receptive field (mrf).

```
function mrf = LS_strf(psth, spectrogram, time)
i = 4; size(spectrogram, 2); % analyzed time window:
    0–30 ms; bin: 10 ms % part of spectrogram (within
    analyzed time window) is converted to single row
spec = [spectrogram(:, i) spectrogram(:, i - 1)
(spectrogram)(:, i - 2) spectrogram(:, i - 3)];
psth = psth(i); % psth
time = time(i); % time course of stimulus
[pc, scores, latent, tsquare] = princomp(spec); %
preprocessing with PCA
lv = fix(0.2 * size(pc, 1)); % taking top 20% of PCs
a = [scores(:, 1:lv) time'] \ psth'; % LS model
% recovering STRF from LS coefficients
mrf_pca = pc(:, 1:lv) * a(1:lv);
sz = size(spectrogram, 1);
for i = 1: size(spec, 2) / sz % no. of time lines of RF
mrf(:, i) = mrf_pca(sz * (i - 1) + 1: sz * i);
end
figure; imagesc(mrf)
```

This function, together with sample data and additional code, can be downloaded from <http://spikelab.jbpierce.org/strf>.

## References

Aertsen AMHJ, Johannesma PIM. The spectro-temporal receptive field: a functional characteristic of auditory neurons. *Biol Cybern* 1981;42:133–43.

Attias H, Schreiner CE. Low-order temporal statistics of natural sounds. In: Mozer MC, Jordan MI, Petsche T, editors. *Advances in neural information processing systems 9*. Cambridge, MA: MIT Press; 1997.

Calhoun B, Schreiner C. Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics. *Eur J Neurosci* 1998;10:926–40.

de Boer R, Kuyper P. Triggered correlation. *IEEE Trans Biomed Eng* 1968;15:169–79.

de Ruyter van Steveninck RR, Bialek W. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc R Soc B* 1988;234:379–414.

DeAngelis GC, Ohzawa I, Freeman RD. Receptive-field dynamics in the central visual pathways. *Trends Neurosci* 1995;18:451–8.

deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. *Science* 1998;280:1439–43.

Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 2001;85:1220–34.

Donoho DL. De-noising by soft-thresholding. *IEEE Trans Inform Theor* 1995;41(3):613–27.

Escabi MA, Schreiner CE. Non-linear spectro-temporal envelope processing in the cat IR. *Assoc Res Otolaryngol Abstr* 1999;22:869.

Hackett TA, Preuss TM, Kaas JH. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 2001;441(3):197–222.

Hyvärinen A. Survey on independent component analysis. *Neural Comput Surv* 1999;2:94–128.

Kelly KA, Metzger RR, Mullette-Gillman OA, Werner-Reiss U, Groh JM. Representation of sound location in the primate brain. In: Ghazanfar A, editor. *Primate audition: behavior and neurobiology*. Boca Raton, FL: CRC Press; 2003.

Lawson CL, Hanson RJ. Solving least-squares problems (series: classics in applied mathematics: 15). Philadelphia: SIAM; 1995.

Lewicki MS. Efficient coding of natural sounds. *Nat Neurosci* 2002;5:356–63.

Liang L, Lu T, Wang X. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol* 2002;87:2237–61.

Middlebrooks JC, Clock AE, Xu L, Green DMA. Panoramic code for sound location by cortical neurons. *Science* 1994;264:842–4.

Middlebrooks JC, Xu L, Clock AE, Green DM. Codes for sound-source location in nontopographic auditory cortex. *J Neurophysiol* 1998;80:863–81.

Quatieri TF. *Discrete-time speech signal processing: principles and practice*. Upper Saddle River, NJ: Prentice Hall, PTR; 2002.

Recanzone GH, Schreiner CE, Sutter ML, Beitel RE, Merzenich MM. Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *J Comp Neurol* 1999;415:460–81.

Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 1995;268:111–4.

Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W. *Spikes: exploring the neural code*. Cambridge, MA: MIT Press; 1997.

Sahani M, Linden JF. How linear are auditory cortical responses? In: Becker S, Thrun S, Obermayer K, editors. *Advance neural information processing system*, vol. 15. Cambridge, MA: MIT Press; 2003.

Salinas E, Sejnowski TJ. Integrate-and-fire models driven by correlated stochastic input. *Neural Comput* 2002;14:2111–55.

Schreiner CE, Mendelson JR, Sutter ML. Functional topography of cat primary auditory cortex: representation of tone intensity. *Exp Brain Res* 1992;92:105–22.

Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 2001;86:1445–58.

Shamma SA, Versnel H, Kowalski N. Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Audiol Neurosci* 1995;1:233–54.

Srikantan NS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM. Representation of spectral and temporal envelope of

- twitter vocalizations in common marmoset primary auditory cortex. *J Neurophysiol* 2002;87:1723–37.
- Theunissen FE, Doupe AJ. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J Neurosci* 1998;18:3786–802.
- Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. *J Neurosci* 2000;20:2315–31.
- Wang X. On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci* 2000;97(22):11843–9.
- Wang X, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 1995;74:2685–706.
- Wickerhauser MV. *Adapted wavelet analysis from theory to software*. A K Peters Ltd.; 1994.
- Willmore B, Smyth D. Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. *Network: Comput Neural Syst* 2003;14:553–77.