College of Engineering

Department of Biomedical Engineering

Senior Project Final Report

Modeling of Spectro-Temporal Receptive Fields and Simulation of the Population Response in Zebra Finch Field-L

Submitted in Partial Fulfillment for the Degree of Bachelor of Science.

21 April, 2006

Revised 29 June 2006

Student:	Graham Voysey
Technical Advisor:	Professor Kamal Sen
Senior Project Advisor:	Professor Kenneth Lutchen

Abstract

The zebra finch is a good model species for the study of complex sound learning and perception. Zebra finch song is spectrotemporally complex, feature-rich, and has similar acoustic properties to human speech. Field L, the avian analogue of auditory cortex, is the area of the brain in which songs are analyzed and processed. Previous studies have quantified the responses of single auditory neurons to sound using Spectro-Temporal Receptive Fields (STRFs). Though a collection of these STRFs exist, they have not been systematically modeled and characterized to show the variety of feature encoding that field L possesses. Knowledge of the diversity of field L could impact studies into neural coding, speech learning, and speech perception. Therefore, this project created models of STRFs, validated these models, used these models to generate firing-rate responses to songs, and characterized these responses visually into the so-called population response.

Modeling was accomplished by a least-squares curve fitting of experimentally characterized STRFs to Gabor Functions. These have been previously shown to model neural data well. Firing rate generation was accomplished by the convolution of model responses to a series of birdsongs. EarLab, a software suite specializing in modeling and visualization of the auditory system, was used to visualize the population response.

The visualizations of the population response revealed the representation of spectrotemporally complex sound by neural population in field L.

This project would not have been possible without the guidance and insight of the members of the Natural Sounds and Neural Coding Lab: Professor Kamal Sen, postdoctoral student Cyrus Billimoria, and graduate students Rajiv Narayan and Gilberto Graña. The EarLab development team, and in particular, David Anderson, made development of the EarLab module possible. The research done by Qiu, DeAngelis, Sen, Theunissen, Doupe, and many others was critical to the completion of this project. The tacit support of DK, LL, LT, RMS, KT, DR, and MK was invaluable.

Contents

List of Figures

1 Introduction

The neural processes that govern the perception of sound in the auditory processing areas of the brain are poorly understood and largely unknown. The response characteristics of the auditory cortex have historically been categorized and studied with low-resolution methods such as MRI, which give qualitative information about the activation or deactivation of different cortical subregions. They fail to provide precise numerical data about the firing rates of individual neurons. In particular, the processes by which the brain interprets complex natural sounds, such as song or speech, remain largely unknown.

The zebra finch (*Taeniopygia guttata*) is a good model organism for studying the cortical processing of complex sounds in humans[6]. Both speech and song are learned by an organism and proper hearing is mandatory for proper learning of both sounds. The Field-L complex of the zebra finch brain is the avian analogue of primary auditory cortex. It has become an area of particular interest because it is thought to play an important role in the perception of complex sounds such as songs[2].

The Spectro-Temporal Receptive Field (STRF) is a function of time and frequency commonly used to represent neuron behavior. Each point on the time-versus-frequency plot takes the value of the neurons' firing rate relative to a stochastic baseline. STRFs can be used as neuronal transfer functions by which output neuronal firing rate for a given complex auditory input can be predicted.

Previous experiments have studied the stimulus-response functions of individual neurons in Field-L to a range of stimulus types[2][6][7]. Due to inherent limitations in the recording techniques in use, no studies that examine the simultaneous response of many neurons in Field-L exist: simultaneous multi-neuronal recordings are prohibitively difficult to obtain. To begin to examine multi-neuron behavior (commonly called the population response), it is necessary to use computational models of many Field-L neurons in a simulation of activity. By reducing complex experimental data to a combination of one or more mathematical functions while ensuring that the model accurately represents the data recorded, a wide array of analytical methods can be brought to bear on the system. No such studies of the population response have yet been performed. The assembly of Field-L neuron models needed for such a study are also absent.

The Gabor function has been shown to be capable of modeling auditory STRF behavior to a very high degree of accuracy[5]. It is therefore a promising candidate for a system of basis functions. Nonlinear least-squares modeling, an algorithm that creates models by minimizing the error between a parametric equation and a set of data, has been used with success in the past to fit STRFs to Gabor functions[5][4].

Auditory simulation software already exists in the form of EarLab, a modular, extensible software package developed at Boston University's Hearing Research Center. It seeks to integrate data from many locations into an accessible, online database for the simulation of all parts of the auditory system. It does not yet have models of Field-L[1].

This project therefore sought to create and validate a set of Gabor-modeled STRFs, to use them to generate firing rate data of many neurons in Field-L, and to visualize and quantify the population response. Knowledge of the overall behavior of Field-L is critical in the understanding of the neural coding that governs perception of spectrotemporally complex, feature rich sounds.

2 Background

2.1 Zebra Finch Field L

In the Zebra Finch, Field L is the earliest area of the auditory system to show selectivity for songs[2]. Field L receives information from the lower auditory system, and provides simultaneous input to the Anterior Forebrain Pathway (AFP) and the Motor Pathway (MP). The AFP has been shown[2] to act as a training feedback system for juvenile birds in the acquisition of birdsong. The MP governs song production in both adult and juvenile birds[2].

Figure 1 shows the relationships between Field L, the AFP, and the MP.

It has been found that Field L responses can be highly nonlinear[6][7][9]. In particular, Solis *et al*[7]

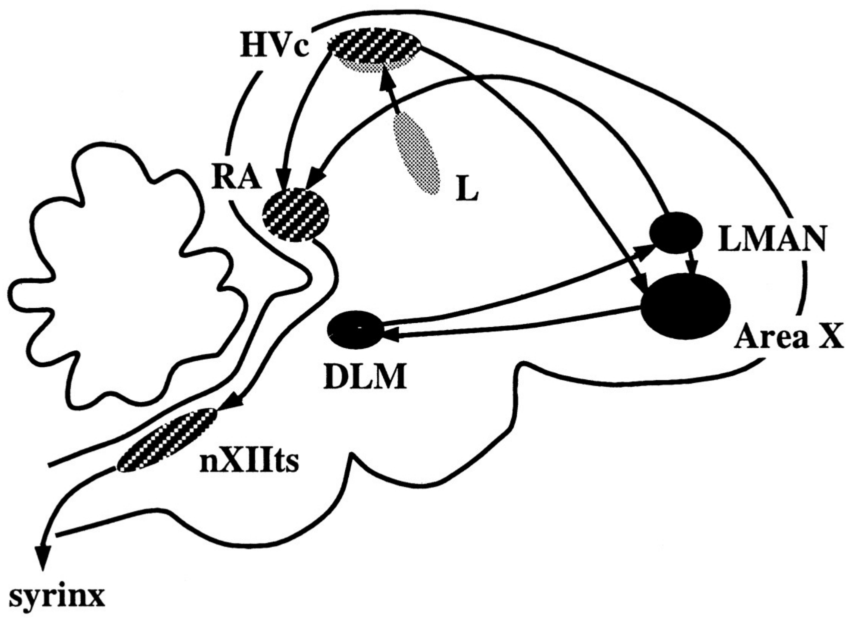

Figure 1: A schematic diagram of the auditory circuitry in zebra finch cortex. Field L, the area of interest in this project, is show in the center. Arrows indicate the innervation pathways from Field L to the AFP, in black, and the MP, in striped. The outlined border shows the physiological location of Field L, the AFP, and the MP with respect to the contours of zebra finch cortex.

demonstrate that the mature zebra finch auditory cortex responds to the complex spectrotemporal patterns in the bird's own song much more strongly than any other input. This makes traditional analytical methods poor choices for characterizing Field-L behavior (see Section 2.3). Theunissen *et al*[9] have pioneered new techniques that remove many of the limitations that hamper traditional methods.

2.2 The Spectro-Temporal Receptive Field

The Spectro-Temporal Receptive Field is a common means of describing the behavior of auditory neurons. It is particularly useful when attempting to characterize the behavior of neurons – such as those in Field L – that exhibit nonlinear responses or respond to specific spectrotemporal cues preferentially. STRFs are discrete functions of time and frequency. The value of the STRF at each time-frequency sample is the firing rate of the neuron being represented.

All neurons have a stochastic baseline firing rate that is nonzero – they fire randomly in the absence of

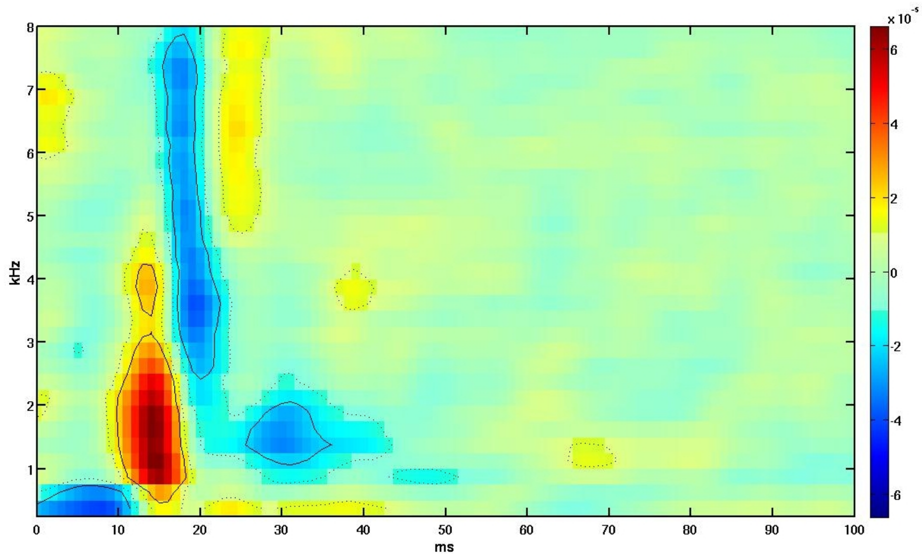

Figure 2: The STRF for neuron `red3-13-oldsongs20`. The response for the first 100ms post-stimulus is plotted. The red region is one of high excitation, and there is a strong, wideband region of inhibition following it in blue. Smaller regions of inhibition are also present.

any stimuli at a low level. A firing rate above threshold is a response to an activating stimulus, and a firing rate that falls below threshold is a response to an inhibitory stimulus. When the STRF is visualized, it is done so as a color-coded time-frequency plot. Regions of activation are shown in “hot” colors, and regions of inhibition are shown in “cold” colors. Figure 2 shows an example STRF. The neuron it models is from Sen *et al.* Their “`oldsong20`” collection of recordings was used extensively in this project.

STRFs in Field-L exhibit a diversity of characteristic responses. A sampling of these are shown in Figure 4. Five distinct STRF classes have been seen in statistical analysis of the STRF populations measured from Field L[8]. 21% are Broadband Onset STRFs, with wide-spectrum activation followed by wide-spectrum deactivation. 5% are Narrowband Onset STRFs, with narrow frequency activation followed by narrow frequency deactivation. 27% have excitatory areas sandwiched between areas of inhibition, 7% have only activation sites, and the remaining 29% have complex patterns of activation and inhibition. STRFs representative of these areas are shown in Figure 3, which is taken from the work of Theunissen *et al*[8].

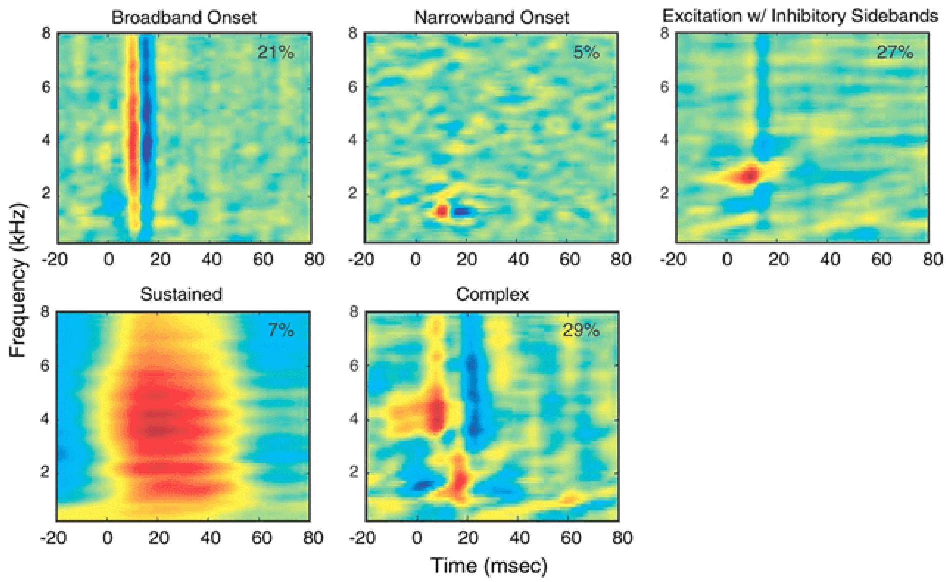

Figure 3: Commonly seen STRF types with associated frequency, as originally published by Theunissen *et al*.

2.3 STRF Creation

The traditional method for obtaining a STRF for a given neuron is reverse correlation, a method of spike-triggered averaging. White noise is used as an input to the neuron many times, and the resulting spike trains are averaged to give the STRF[9]. For linear neurons, the STRF is exactly the transfer function of that neuron, and for nearly-linear neurons, it is a very good approximation. This makes the STRF ideal for calculating responses to a wide range of stimuli.

In high levels of the auditory system, reverse correlation with a white noise stimulus is a poor method of obtaining the STRF, because the full dynamic range will not be stimulated. Instead, they respond preferentially to spectrotemporal features that, by definition, white noise does not possess. The use of natural sounds that contain these features to obtain STRFs was pioneered by Theunissen *et al*[9]. Their method uses a numerical technique to extract the STRF from the response to birdsong. Experimentally gathered STRFs were decomposed with the Singular Value Decomposition (SVD) technique into a linear sum of time-frequency separable STRF components.

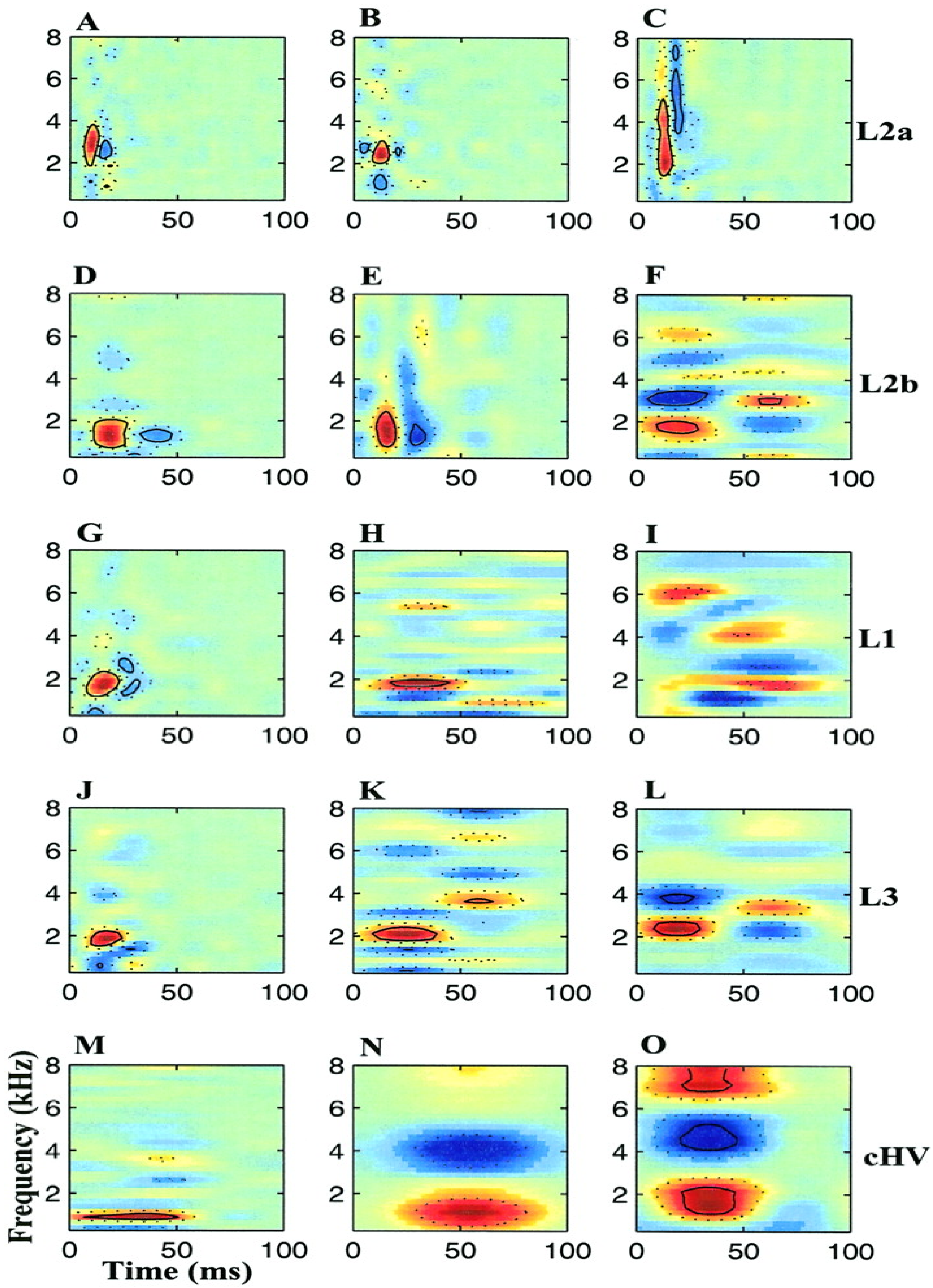

Figure 4: A wide spectrum of STRFs found in Field-L are shown. The time axes are the first 100ms of response, and the frequency axis spans 8kHz of response. The labels on the right hand side indicate which region of Field-L these STRFs originate from. This figure was originally published by Sen *et al*[6].

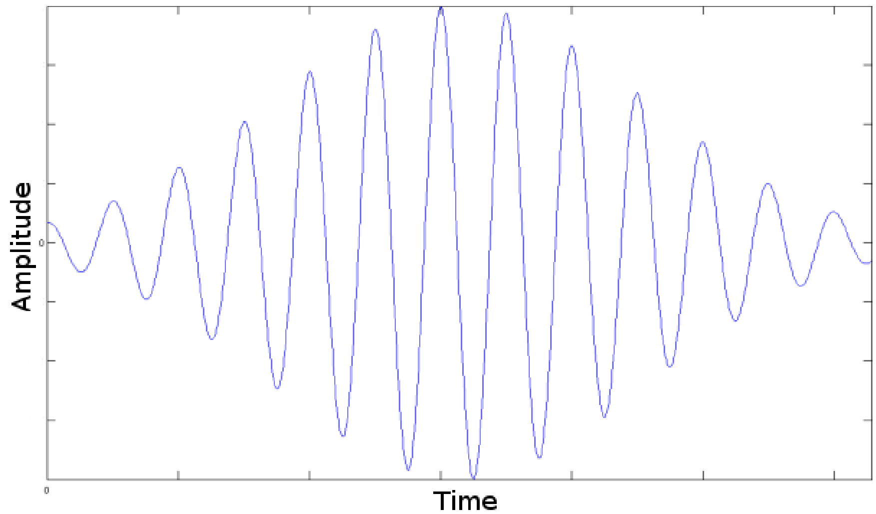

Figure 5: A sample Gabor function. It is clear that Gabor functions are amplitude modulated sinusoids with Gaussian envelope.

2.4 Modeling Data: Applicability of the Gabor Function

The need for a means to model STRFs is met by the Gabor function. Qiu *et al*[5] present a method in which every STRF component obtained by SVD is modeled by two Gabor functions: one representing the time waveform and one representing the frequency waveform.

The Gabor function $G(x)$, in general, takes the form:

$$G(x) = \alpha e^{-\left(\frac{2(x-x_0)}{\beta}\right)^2} \cos\left[2\pi\gamma(x - x_0) + \epsilon\right] \tag{1}$$

where α, β, γ, and ϵ are free parameters. Figure 5 shows a sample Gabor function. Conceptually, a Gabor function is an amplitude-modulated sinusoid whose envelope function is a Gaussian. An important feature of the Gabor function is that its free parameters correspond to physiologically meaningful neural characteristics. This is covered in more detail in Section 3.3.4.

Gabor functions form a powerful set of basis functions for modeling STRFs, because several Gabor functions can be combined to fit the patterns of excitation and inhibition that are characteristically seen in STRFs in general, and in STRFs measured from Field-L in particular (see Figure 2).

2.5 Estimation of Firing Rate

Because STRFs are transfer functions, they are used to compute the firing rate of the neuron they represent. The input to a STRF is a song spectrogram, which is a time-frequency plot of a song.

$$SPEC(t, \omega) * STRF(t, \omega) = FR(t) \tag{2}$$

The output from a STRF is a firing rate pattern for the neuron in response to a song.

2.6 The EarLab Computational Environment

EarLab is a suite of auditory modeling software developed at the Boston University Hearing Research Center. The three main programs within EarLab are the Desktop Modeling Environment, the Distributed Modeling Environment, and the EarLab Data Viewer. The Data Viewer is of particular applicability to this project as a visualizer of many simultaneous firing rates. Single firing rate responses can be extracted easily, and the visualizations can be scaled, zoomed, and otherwise manipulated to easily show the complex features of concerted firing rate data.

Ultimately, the Distributed and Desktop Modeling Environments will be of use to this project. They provide a means by which modeling and computation of many kinds of auditory data can be accomplished. For extraordinarily large data sets, the distributed environment is particularly useful. It can use many separate computers over a network to parallelize the processing of complex data. The desktop modeling environment is limited to just one computer.

2.7 Current Challenges

Gabor-modeled STRFs have been studied in auditory neurons in cats, at a pre-cortical level of the auditory system[5].

The current methods for STRF modeling are limited in several key areas.

1. Qiu's work on Gabor models only modeled near-linear STRFs in cat inferior colliculus, and had mixed results in the goodness of fit[5]. At the beginning of this project, it was unclear that the Gabor formalism could account for STRF structure in Field-L and primary auditory cortex.

2. There is no centralized bank of many STRFs for Field-L and primary auditory cortex – consequently, while the response characteristics of one neuron may be studied, there exists no easy way to study the population response of many neurons acting in concert.

3. The lack of EarLab modules for Field-L restricts the types of modeling that can be performed and the ease of visualization.

3 Methodology

3.1 Overview

The methodology presented in this section has been adapted from the work of Qiu *et al*[5] and Theunissen *et al*[9]. Qiu developed a Gabor-based modeling method for neurons in cat inferior colliculus, and Theunissen showed that the same approach can be successfully used to model neurons elsewhere.

The methodology consists of several discrete phases. First, data must be recorded from a zebra finch to acquire the STRF. This is not within the scope of this project, but is mentioned in brief for completeness. Second, the recorded STRF is digitally stored and discretized. Each discrete component is then modeled with two Gabor functions. These component models are then weighted with respect to their contribution to the shape of the STRF, and reassembled into the overall model. This model is validated with several goodness-of-fit metrics. After each model is generated, it is then arranged such that EarLab can incorporate it as a module and use it to simulate Field-L population responses.

3.2 Acquisition of Experimental STRFs

The STRFs modeled in this project were previously recorded and digitally stored following the general procedure given by Theunissen *et al*[9]. In brief, all measurements were taken from adult anesthetized zebra finches by a tungsten electrode inserted through the neostratium, the area of the zebra finch brain where Field-L is located. Appropriate electrode placement into Field-L was verified histologically post-experimentation. The sampling rate was 1 KHz over a total time window of 600ms and a frequency range of 250 Hz to 8 KHz with a step size of 250 Hz. The experimental STRFs are therefore 31-row, 601-column matrices. The stimulus was applied 300ms into the recording. The experimental STRFs used to create models result from the song recordings in Sen *et al*'s `oldsongs20` set.

3.3 Model Generation

The generation of STRF models was accomplished by the creation of `gabor_strf`, a collection of MATLAB scripts that was written to take experimental STRFs and create models of them. The details of `gabor_strf` are presented in Section 4.1; a source code snapshot can be found in Appendix 1. The methodology presented below is formulated algorithmically; details about how `gabor_strf` specifically implements the following sections can be found in Section 4.

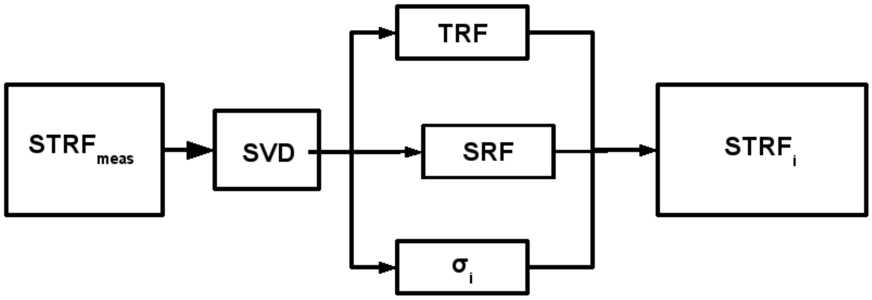

Figure 6: A block diagram of SVD. The experimental STRF is decomposed into a Temporal Receptive Field, a Spectral Receptive Field, and a vector of singular values. Elements of the TRF, SRF, and the singular values are combined to create many component STRFs.

3.3.1 Separation of Experimental STRF

Singular Value Decomposition is applied to the experimental STRF.

$$STRF_{meas} = SRF \times \sigma \times TRF^{*} \tag{3}$$

The experimental STRF is decomposed into three constituent matrices. The transposition operator is represented by $*$. Formally, it is a Hermetian transpose, but this is irrelevant for data entirely in $\Re$. SRF is the Spectral Receptive Field, a square matrix whose dimension corresponds to the number of frequency channels used to record $STRF_{meas}$. It represents the frequency-dependent behavior of the neuron. σ is a diagonal, non-square matrix of like dimension to $STRF_{meas}$. The values of σ are called *singular values*. TRF is the Temporal Receptive Field, a square matrix whose dimension corresponds to the number of time samples in $STRF_{meas}$. It represents the time-dependent behavior of the neuron. Figure 6 shows a block diagram of this process.

3.3.2 Creation of Component STRFs

The matrices that result from the decomposition of the experimental STRF represent a series of STRF components which are linearly-independent subsets of the overall STRF. The i^{th} component of the overall

STRF, therefore, is given as

$$STRF_i = \sigma_i \times \mathrm{srf}_i \times \mathrm{trf}_i^* \tag{4}$$

where srf_i is the i^{th} column of SRF and trf_i is the i^{th} row of TRF from equation (3). As the value of i increases, the magnitude of the associated singular value decreases and the contribution of $STRF_i$ to the overall STRF is attenuated.

3.3.3 Noise Determination by SVD

An important consequence of this attenuation is that it provides a useful means of filtering noise. Those portions of the STRF that represent stochastic firing that do not depend on the stimulus should not be modeled. Because the absolute value of the stochastic firing rate is small compared to the firing rates of highly excitatory or inhibitory areas, contributions to the overall STRF by stochastic firing rates will be associated with small-valued singular values. Because the system is causal, any firing of the neuron in the 300ms prior to stimulus onset is not due to the neuron responding to the spectrotemporal features of the stimulus. Therefore, the maximum singular value obtained by the decomposition of the first half of the STRF is used to threshold the singular values of the whole STRF. Any singular value of the entire experimental STRF that falls below the maximum singular value of the acausal region of the experimental STRF represents noise and can be disregarded. This method of noise determination is conservative. It places more importance on capturing features characteristics than minimizing noise. This has some consequences for fit values. For more detail, see Section 5.

This is demonstrated in Figure 7, which is taken from the neuron `ae-14-oldsongs20`.

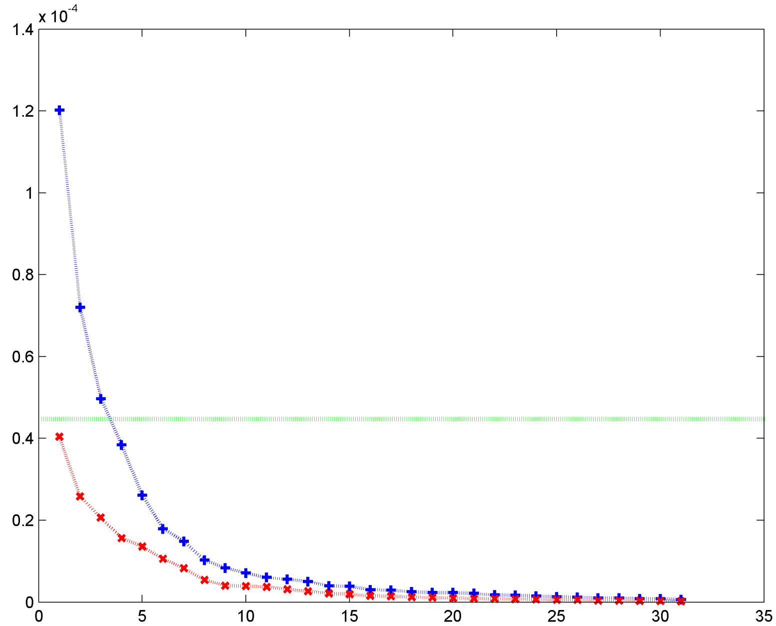

Figure 7: Singular Values as a noise filtering mechanism. The value of each singular value is plotted versus its rank index for both the entire STRF, in blue, and the acausal portion, in red. Those singular values of the overall STRF whose values exceed the value of the maximum acausal singular value contribute significantly to the STRF, and are above the green threshhold line.

3.3.4 Component Gabor Fitting

Each component of TRF and SRF that correspond to those singular values found to be above the noise threshold are fit to Gabor functions by a least-squares algorithm. This algorithm iteratively fits a parameterized equation to data such that the mean squared error between the equation and the data is minimized[4]. Figure 8 shows this process as a block diagram.

The theory behind least-squares fitting is well understood and applicable in many fields. In brief, it provides a method to produce a mathematical function that best mimics the behavior of discrete experimental data. In this case, the basis function that is used by the least-squares algorithm is the Gabor function, and the merit function is the mean squared error between the experimental data and the Gabor function, as given in Equations 6 and 8.

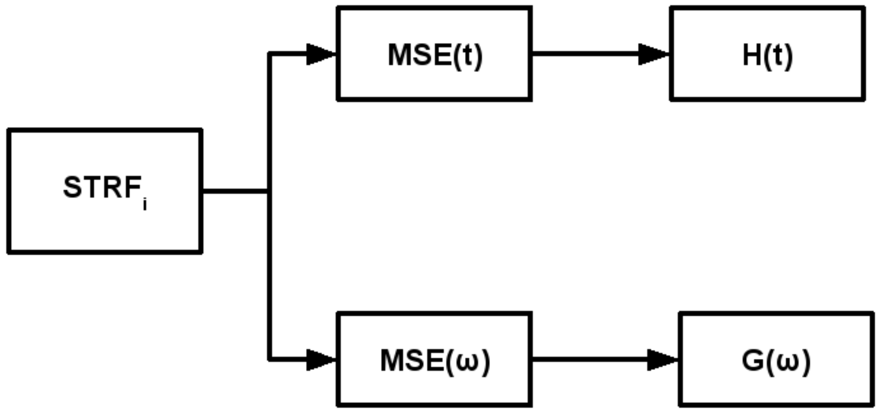

Figure 8: STRF Modeling. Each STRF component is fit by minimizing the Mean Squared Error to individual Gabor functions, H and G

The algorithm is given some initial starting conditions for values of each of the parameters, and it iteratively modifies these values until the merit function's change between each iteration hits some minimal threshold.

The SRF is fit to $G_i(\omega)$:

$$G_i(\omega) = Ke^{-\left(\frac{2(\omega-\omega_0)}{BW}\right)^2} \cos\left[2\pi\Omega_o(\omega-\omega_0)+P\right] \tag{5}$$

K is the magnitude of the response measured in $\frac{\text{spikes}}{\text{dB}\cdot\text{s}}$, ω_0 is the center frequency of the response in dB Hz, BW is the bandwidth of the enveloping Gaussian, Ω_0 is the best ripple density in $\frac{\text{Hz}}{\text{octave}}$, and P is the phase with respect to ω_0.

Mean squared error between $G_i(\omega)$ and the SRF is given as

$$MSE = \sum_{\omega}\left(\frac{SRF_i - G_i(\omega)}{SRF_i}\right)^2 \tag{6}$$

The temporal profile $H_i(t)$ is similarly fit:

$$H_i(t) = Ke^{-\left(\frac{2(t-t_0)}{D}\right)^2} \cos\left[2\pi F_{m_0}(t-t_0)+Q\right] \tag{7}$$

Here, K is the magnitude of the response measured in $\frac{\text{spikes}}{\text{dB}\cdot\text{s}}$, t_0 is the peak latency, D is the response duration, F_{m_0} is the best modulation frequency, and Q is the phase of the sinusoidal carrier relative to t_0.

Mean squared error between $H_i(t)$ and the TRF is given as

$$MSE = \sum_t \left(\frac{TRF_i - H_i(t)}{TRF_i}\right)^2 \tag{8}$$

3.3.5 Model STRF Assembly

With each significant decomposition of $STRF_{meas}$ now modeled by two Gabor functions, $H_i(t)$ and $G_i(\omega)$, a model STRF is assembled:

$$STRF_m(t,\omega) = \sum_{i=1}^{k} K_i \cdot G_i(\omega) \cdot H_i(t) \tag{9}$$

$STRF_{meas}$ is the linear sum of the component model STRFs. Only the first k singular values are used to reassemble the model. The value of k is determined as the number of singular values that lie above the noise threshold, as described in Section 3.3.3. Typically, $k \leq 4$.

3.4 Firing Rate Estimation

The assembled model STRF can now be used as a transfer function to calculate the firing rate of a neuron in response to a song. The song must first be transformed from a pressure waveform into a spectrogram, and then convolved with the model STRF.

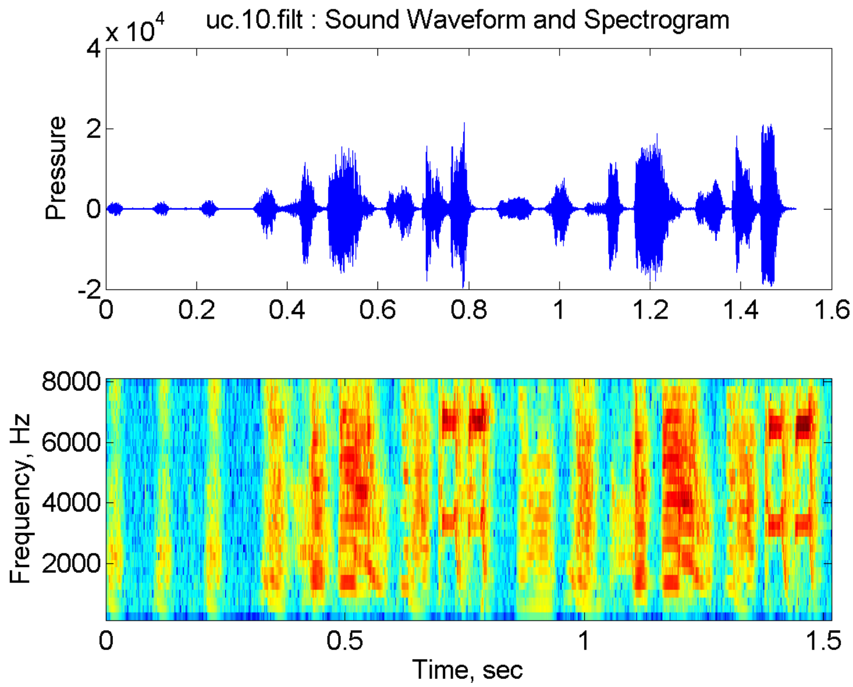

Figure 9: A song, represented as both a time-pressure plot (top) and a time-frequency spectrogram (bottom).

3.4.1 Song Spectrogram Construction

Song recordings exist as a time-pressure waveform. In order to be convolved with the STRF to generate an output, they must be converted to spectrograms – time-frequency plots that represent the energy of the frequency content of a signal over time. The song spectrogram is the magnitude of the Fast Fourier Transform of a time-amplitude plot in a localized moving time window. The song `uc.10.filt` is shown in Figure 9 as both a time-pressure waveform and as a spectrogram.

3.4.2 Firing Rate Generation

The spectrogram is convolved with the STRF to give a firing rate.

$$FR = Spec * STRF \tag{10}$$

The firing rate is then scaled to appropriate levels by normalizing it against firing rates generated from experimental data from the same cell and the same song.

3.5 Model Validation

For a model to be valid, it must be shown to accurately represent the data it seeks to model. For models of Field L STRFs, there are several means by which this can be accomplished. Each are ways of statistically comparing how similar the model STRF is to the experimental STRF.

3.5.1 Temporal Similarity Index

The temporal similarity index is a cross-correlation between the TRF and $H(t)$:

$$SI_t = \frac{\langle TRF, H(t) \rangle}{\| TRF \| \cdot \| H(t) \|} \tag{11}$$

The $\langle , \rangle$ operator denotes vector correlation, and the $\| \cdot \|$ operator denotes the vector norm. SI_t takes a value between -1 and 1. A value of 1 indicates that the two waveforms are identical, a value of -1 indicates the waveforms are identical but differ by sign, and a value of 0 indicates that the waveforms have nothing in common.

3.5.2 Spectral Similarity Index

The spectral similarity index is also a cross-correlation, between the SRF and $G(\omega)$:

$$SI_\omega = \frac{\langle SRF, G(\omega) \rangle}{\| SRF \| \cdot \| G(\omega) \|} \tag{12}$$

It also assumes a value between -1 and 1.

3.5.3 Mean Squared Error

The mean squared error is a measure of energetic difference between two waveforms. It can be used to characterize the fits for the temporal waveform, the spectral waveform, and the overall STRF:

$$MSE = \frac{\sum_{\omega,t} (STRF_{actual} - STRF_{model})^2}{\sum_{\omega,t} STRF_{actual}^2} \tag{13}$$

$STRF_{actual}$ and $STRF_{model}$ can represent the overall STRF, the SRF, or the TRF: all three comparisons are meaningful. Values are always positive, with lower values indicating better fits.

3.6 Limitations of Separability

An inseparable STRF is one for which the first singular value will contain a vast majority of the nature of the STRF, and very little of significance is relegated to lower-order singular values. This limits the maximum fit possible. A separability index, measuring how completely separable a given $STRF_{meas}$ is, is computed using σ_i, the singular values:

$$\alpha_d = \frac{\sigma_1^2 - \sum_{i=2}^{N} \sigma_i^2}{\sigma_1^2 + \sum_{i=2}^{N} \sigma_i^2} \tag{14}$$

Values of α_d range from 0 to 1, with values near 1 indicating a well-separable $STRF_{meas}$. The separability of the spectral and temporal characteristics of a STRF is a property of the STRF that reflects the dynamics of the neuron it models.

As there is considerable interest in modeling even strongly inseparable STRFs, it is important to note that the SVD procedure performed in this modeling algorithm performs comparably well in cells that are separable or inseparable.

3.7 Population Response Visualization

The population response is visualized as a colorized plot of many firing rates from many STRFs, all of which are calculated from the same song spectrogram input. Each firing rate is plotted as a horizontal slice on a neuron-versus-time plot, and colorized like the STRF to indicate high-firing and low-firing areas. This is accomplished with EarLab Data Viewer. Additionally, the responses are plotted simultaneously on a time-amplitude plot.

A primary function of viewing the population response is to more easily note inter-neuron patterns of behavior. As such, means to rearrange plotted responses by various metrics are critical, so that the responses of cells with like metrics (or disparate ones), may be compared easily. Two important and easily obtainable metrics are center frequency (CF) and temporal bandwidth (TBW). The CF of a neuron is the frequency at the site of the response of highest magnitude. The TBW is the time bandwidth at the site of the response of highest magnitude. Normally, this response is excitatory.

3.7.1 EarLab File Specifications

EarLab data is represented by two files. One is a binary file of column-major floating point values. The other is an XML stylesheet that defines the appropriate axes, scale factors, and ranges for visualization.

4 Results

4.1 Gabor-STRF

`gabor-strf` is a collection of MATLAB functions that was written to implement the modeling and validation methods specified in Section 3. The flowchart in Figure 10 shows the flow of data in the program. The program takes experimental STRFs as inputs and outputs modeled STRFs and a host of analytical data about them.

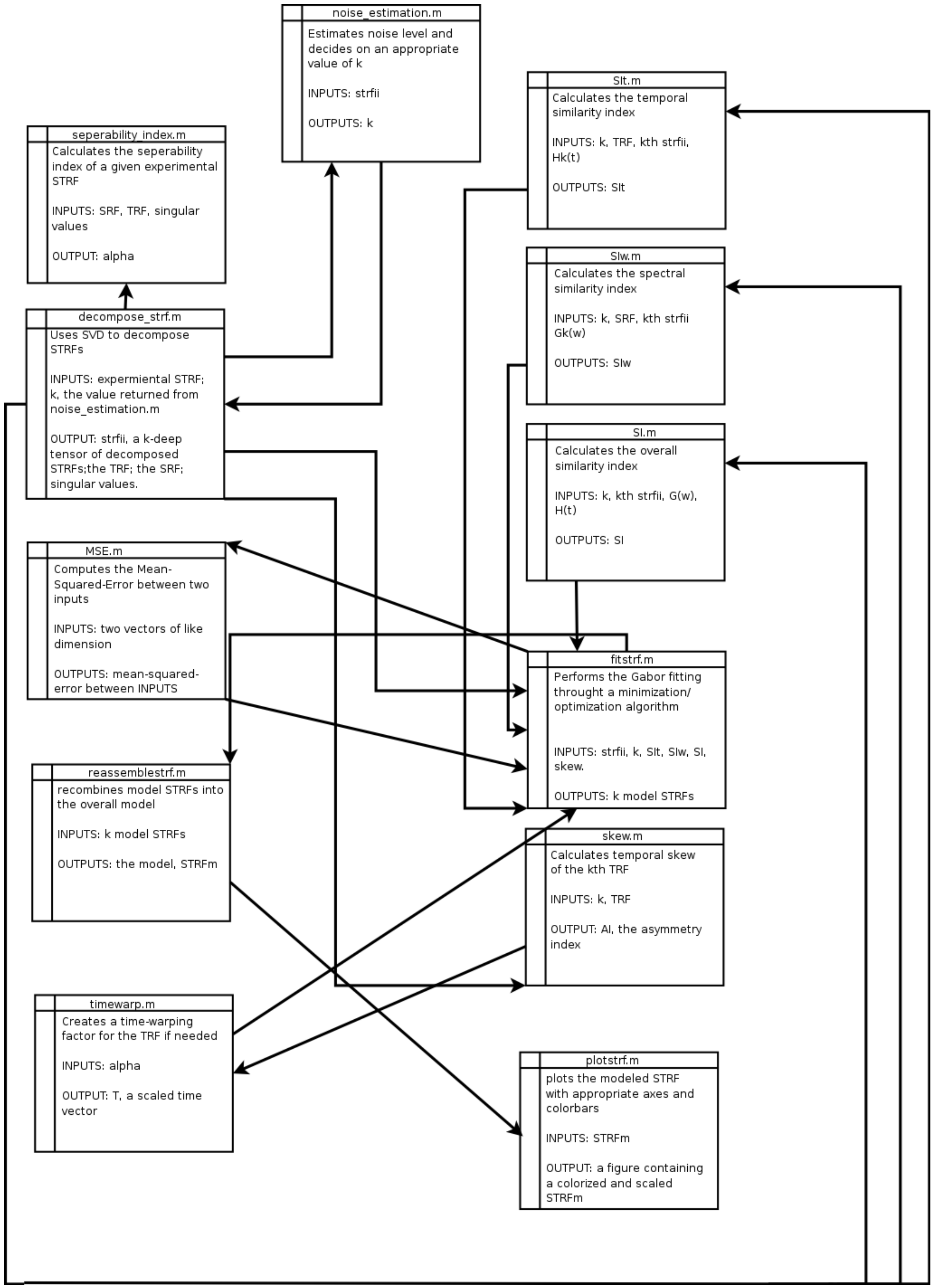

Figure 10: Flowchart of gabor-strf Layout

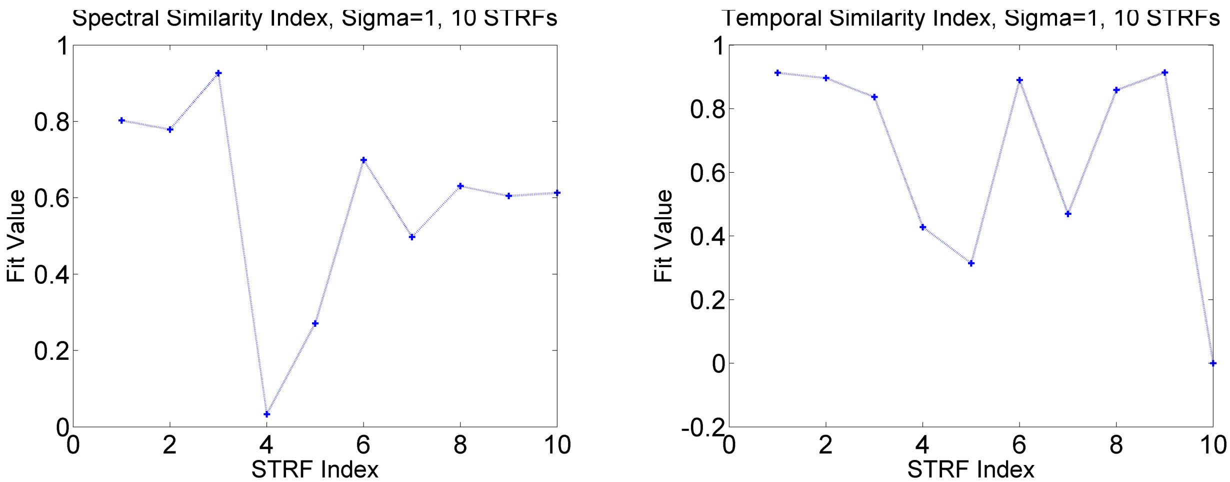

Figure 11: The similarity indices of 10 representative STRFs are plotted. The SI value measures the fit of the first singular value. Fit values are comparable with fits obtained in prior studies.

A copy of `gabor-strf` is included in Appendix I. The source code and a dynamically generated API are also available on the Natural Sounds and Neural Coding Lab's Subversion repository, located at `http://shadowfax.bu.edu/svn/software/gabor_strf`. Subversion is a freely available version control system for the management of large software projects. The current version as of this writing is revision 358. It requires MATLAB v 7.1 or higher with the Optimization toolbox.

4.2 Statistical Fit of Models

4.2.1 Goodness-of-Fit values

Representative data from 10 modeled STRFs are presented in Figure 11. The values for the Similarity Indices in both time and frequency for the first singular value for each STRF are shown. Fits were generally good, and comparable to prior studies[6]. The spectral outlier in STRF 4, and temporal outlier in STRF 10, are probably due to initial parameter sensitivity of the least-squares algorithm. This is discussed in Section 5.2.2. Fits are certainly good enough to use in the visualization of the population response.

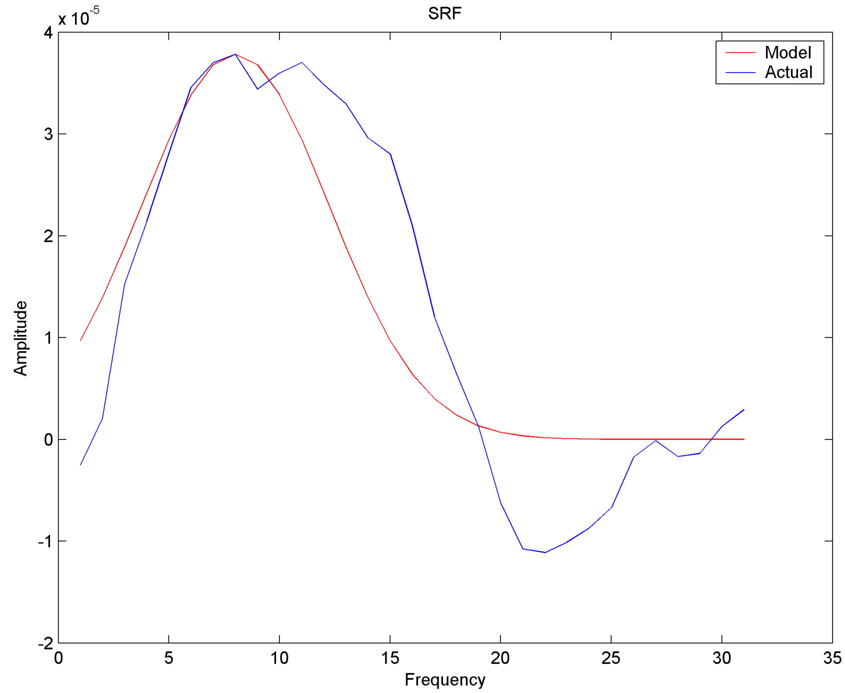

Figure 12: The fit of the first singular spectral receptive field is shown against its model. The model strongly accounts for the initial onset, and has an overall SI of 0.9264

4.2.2 Fits of the Spectral Receptive Field

In Figure 12, the Spectral Receptive Field of the first singular value of `ae-14-oldsongs20` is plotted against its model. This fit had a SI_ω of 0.9264. The model accounts for the peak frequency of activation, but does not give as broad of a response as the experimental data. The model also attenuates the higher-frequency response, which results some loss of fine-structure detail. See Section 5 for further discussion of this fine-structure loss.

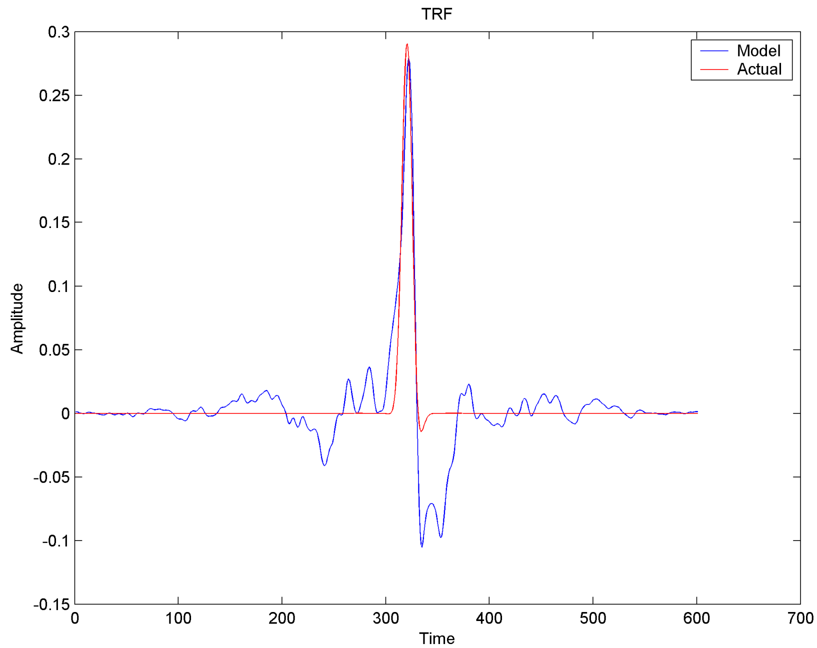

Figure 13: The fit of the first singular temporal receptive field is shown against its model. The model disregards low-amplitude fluxuations, but reproduces the strong excitatory peak with high fidelity. The SI is 0.8366

4.2.3 Fits of the Temporal Receptive Field

In Figure 13, the Temporal Receptive Field of the first singular value of `ae-14-oldsongs20` is plotted against its model. This fit had a SI_t of 0.8366. The model accounts for the strong activation spike and attenuates much of the lower-amplitude response.

4.2.4 Example STRF Models

A wide variety of STRF types are present in Field L. A sampling of modeled STRFs are shown in Figure 14. Some STRFs have only a single activating time-frequency combination, some have many tight bands, and some are wideband inhibitors or wideband excitors. `gabor-strf` can model the diverse STRF types seen in Field L, as discussed in Section 2.2. Each of these firing rates become one STRF's contribution to the population response.

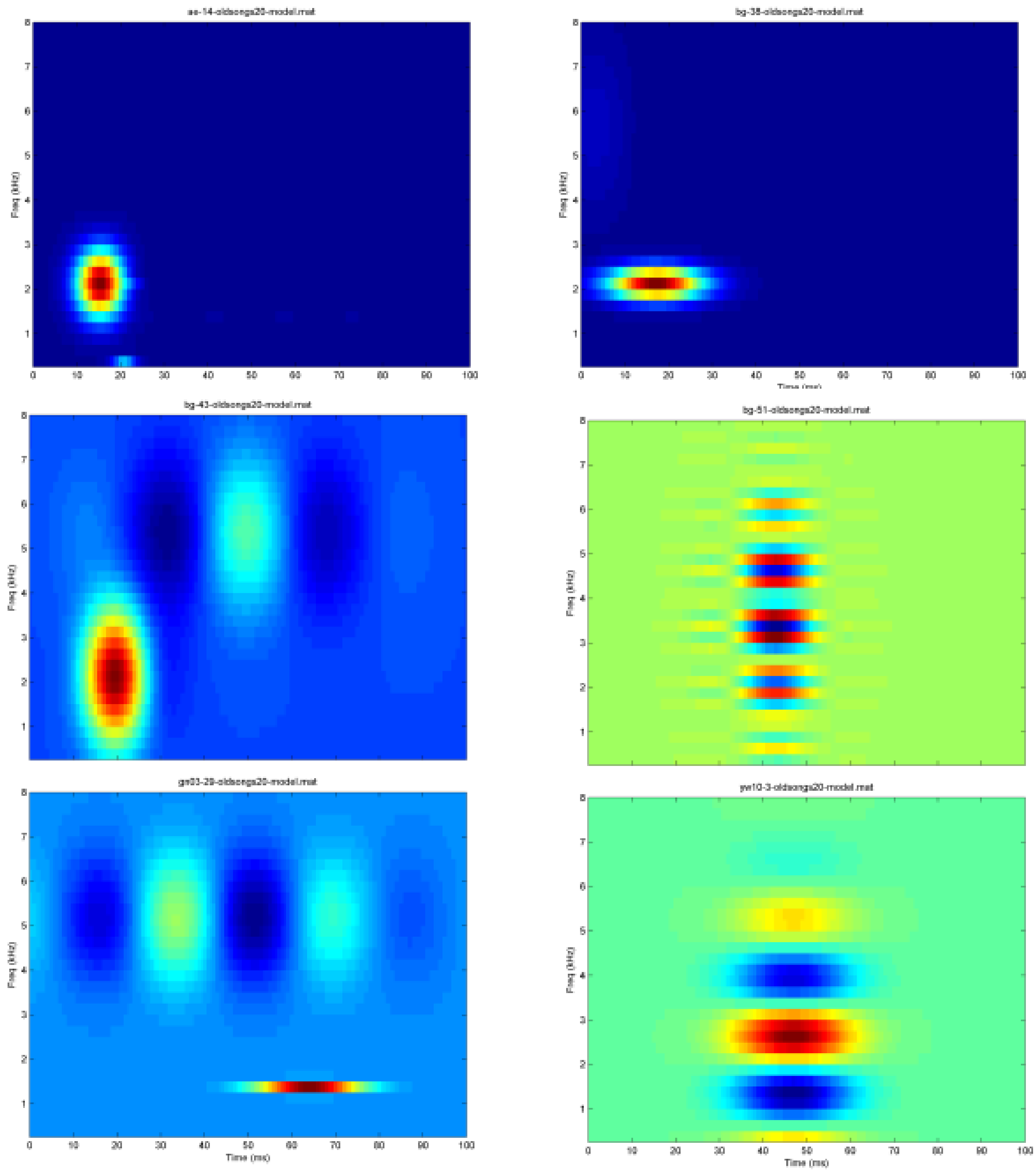

Figure 14: Diverse STRF models from STRFs of Field L. In all cases, time is plotted from 0 to 100ms post-stimulus on the x-axis, and frequency is plotted from 250 Hz to 8 Khz on the y-axis.

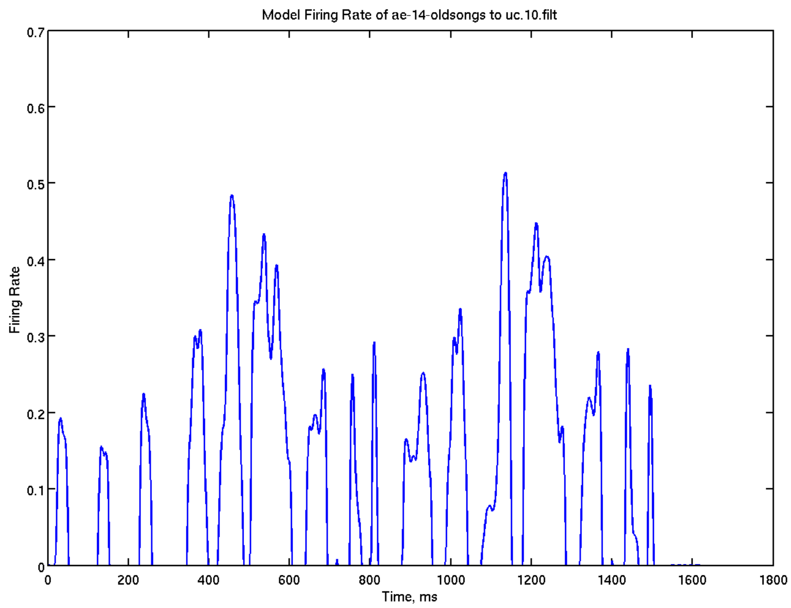

Figure 15: The modeled firing rate of the STRF `bg38-oldsongs20` to the song spectrogram `uc.10.filt`.

4.2.5 Example Firing Rate Outputs

Figure 15 show the output of a STRF model as it is convolved with a song spectrogram. The plot displays the response of one neural model to one birdsong, measured in spikes per second and obtained as described in Section 3.4. The spike bandwidth reflects the slight loss in fine structure from the modeling process.

4.3 Population Responses

EarLab's Data Viewer displays the population response as a sequence of horizontal slices of data. Each slice represents the firing pattern of one neuron. Individual firing patterns are built up on the y-axis in Figure 16. Synchrony, or simultaneous behavior across neurons, is represented as a coherent vertical unit of like color, which indicates that multiple neurons are spiking at the same time. Asynchrony, or lack of simultaneous behavior, is the lack of such vertical correlation.

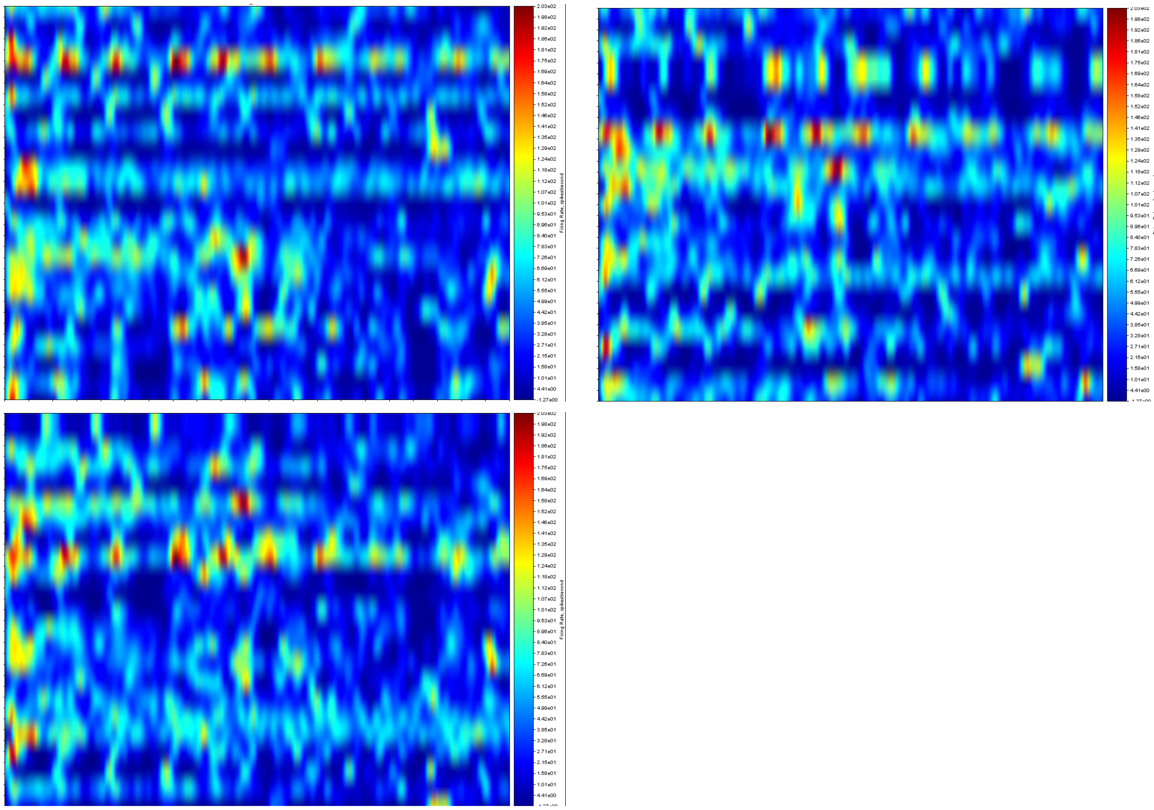

Figure 16: The population response of 10 STRF models to a single song. The same response is sorted in three ways. Top Left: the responses are randomly sorted. Top Right: the response are sorted by CF. Bottom Left: the responses are sorted by TBW. Like STRFs, areas with activating firing rates are colored in hot colors. Each horizontal band represents the firing rate of one neuron. Time is plotted on the x-axis in ms, and the y-axis shows the index of the neuron being modeled.

Correlation between synchronous behavior and CF or TBW (see section 3.7) can be searched for with EarLab. By sorting the order in which the neurons are displayed, possibly correlated neurons are more easily visible. This is also shown in Figure 16.

5 Discussion

5.1 Consequences of Gabor-STRF

The immediate consequence of `gabor-strf` is a clear validation of the applicability to Gabor modeling in Field L. Fit values are good for the first two singular values of all STRFs modeled, suggesting that Gabor functions form an appropriate set of basis functions for modeling STRFs in zebra finch Field-L.

This had not been clear at the beginning of the project. While Gabor functions have been used successfully in prior studies to model neurons, those studies were of lower-order neurons in a different species of model organism[5]. Fit values for Field-L STRFs were good, and nominally better than fits obtained in prior studies[6].

Perfect accuracy of fit values was not a primary goal of this project, as the models are not especially intended for quantitative use. Instead, the qualitative ability of these models to show synchronicity and neuron-neuron interactions in the population response is a more powerful use of `gabor-strf`. The preservation of the envelope shape of the predicted firing rates is therefore more important than retention of the fine structure, and this is what is seen in the models.

5.2 Shortcomings

5.2.1 Theoretical Limitations of Modeling Algorithm

The experimentally measured STRF is a linear model of a known-nonlinear system. It therefore will not perfectly represent the response of a neuron. Statistically fit models of experimental STRFs, as created by `gabor-strf`, therefore will also not perfectly represent the response. In particular, when examining the model firing rate plots simultaneously with the experimental firing rate, it is clear that some fine temporal structure is lost, as seen in Figure 15. However, the model is synchronous to the data, and captures the enveloping waveform of the firing rate very well.

5.2.2 Gabor-STRF

In its current implementation, `gabor-strf` has several drawbacks.

1. The MATLAB function `lsqcurvefit`, which is responsible for performing the least-squares algorithm to fit the data to Gabor functions, is part of the Optimizations Toolbox. Users wishing to run `gabor-strf` must have this toolbox available to them.

2. There is a bug in `lsqcurvefit`. It assumes that data of type `double` is a 64-bit signed floating point value. This is only true on 32-bit architectures. On 64-bit architectures, which are becoming increasingly common and are more powerful and faster, `double`s are assigned a 128-bit value. Consequently, when `lsqcurvefit` is executed on a 64-bit machine, it will not successfully operate because it does not recognize 128-bit numbers as being of the proper type. This restricts `gabor-strf` to running on slower 32-bit machines until The Mathworks resolves this issue in a future patch, or a custom curve-fitting solution is written.

3. There has been a change in how MATLAB handles assignment such that `gabor-strf` will only run on MATLAB version 7 or higher.

4. EarLab Data Viewer must be run on Microsoft Windows operating systems, which removes cross-platform capability for visualization.

5. The sorting algorithm to display population responses sorted by different STRF characteristics is excessively user-intensive and should be improved.

6. The model doesn't account for the slight temporal skew in neuron onset/offset times. This skew is correlated to the physiological refraction time of neurons.

7. `lsqcurvefit` can be sensitive to the initial values assigned to the parameters in the Gabor basis functions.

6 Summary and Recommendations

6.1 Findings

The goals of this project were to determine whether or not the Gabor modeling formalism would be able to model the wide variety of STRF types found in Field L, and to devise a means of visualizing the population response of Field L.

It is clear that the Gabor formalism succeeds. Modeled STRFs have overall good fits and output firing rates that closely resemble actual data.

The visualization of Field L's population response has also been successful. The EarLab framework proved extremely verstaile in its ability to display many simultaneous outputs from STRF models, and has provided a convienent way of noticing synchronous behavior among the neurons of Field L. The proportion and strength of this synchronicity will aid in the understanding of the song processing in Field L, which informs studies of song and speech perception, discrimination, and learning.

6.2 Future Work

A main goal for future research is to re-implement `gabor-strf` in a more portable language. This language should have powerful, tested, built-in math libraries that don't suffer from the drawbacks listed in Section 5.2.2. This would increase flexibility, allow users without MATLAB to use `gabor-strf`, and decrease run times.

A tighter integration with the EarLab framework is also a direction that `gabor-strf` should be taken in. This would allow an even faster analysis time. It would also allow the integration of Field L models into larger-context simulation studies. For example, the inputs to the STRF models are currently birdsongs as heard at the ear - the effects of the lower auditory system are disregarded. EarLab makes modulating these songs with other models representing the lower auditory system a very simple exercise.

A more user-friendly interface to EarLab Data Viewer would be helpful. A tighter integration with constituent data sets could prove useful, instead of a static image-generation modality.

The population response imaging concept is perfectly suited for use in a study of STRF varieties in Field L, particularly one in which the relationship between neuron coactivation and STRF type is investigated.

References

[1] Boston University Hearing Research Center. Earlab: http://earlab.bu.edu. Accessed 01 December, 2005.

[2] A. J. Doupe. Song- and order- selective neurons in the songbird anterior forebrain and their emergence during vocal development. *Journal of Neuroscience*, 17:1147–1167, 1997.

[3] C.J.B. Moore. *Introduction to the Psychology of Hearing*. Academic Press, Boston, MA, fifth edition, 2003.

[4] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipies in C: The Art of Scientific Computing*. Cambridge University Press, 2 edition, 1992.

[5] A. Qiu, C.E. Schreiner, and M.A. Escabí. Gabor analysis of auditory midbrain receptive fields: Spectro-temporal and binaural composition. *Journal of Neurophysiology*, 90:456–476, 2003.

[6] K. Sen, F.E. Theunissen, and A.J. Doupe. Feature analysis of natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology*, 86:1445–1458, 2001.

[7] M.M. Solis, M.S. Brainard, N.A. Hessler, and A.J. Doupe. Song selectivity and sensorimotor signals in vocal learning and production. *Proceedings of the National Academy of Sciences*, 97:11836–11842, 2000.

[8] F.E. Theunissen, N. Amin, S. S. Shaevitz, S. M. N. Woolley, T. Fremouw, and M. E. Hauber. Song selectivity in the song system and in the auditory forebrain. *Annals of the New York Academy of Sciences*, 1016:223–245, 2004.

[9] F.E. Theunissen, K. Sen, and A.J. Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience*, 20:2315–2331, 2000.

APPENDIX I : GABOR - STRF CODE LISTING