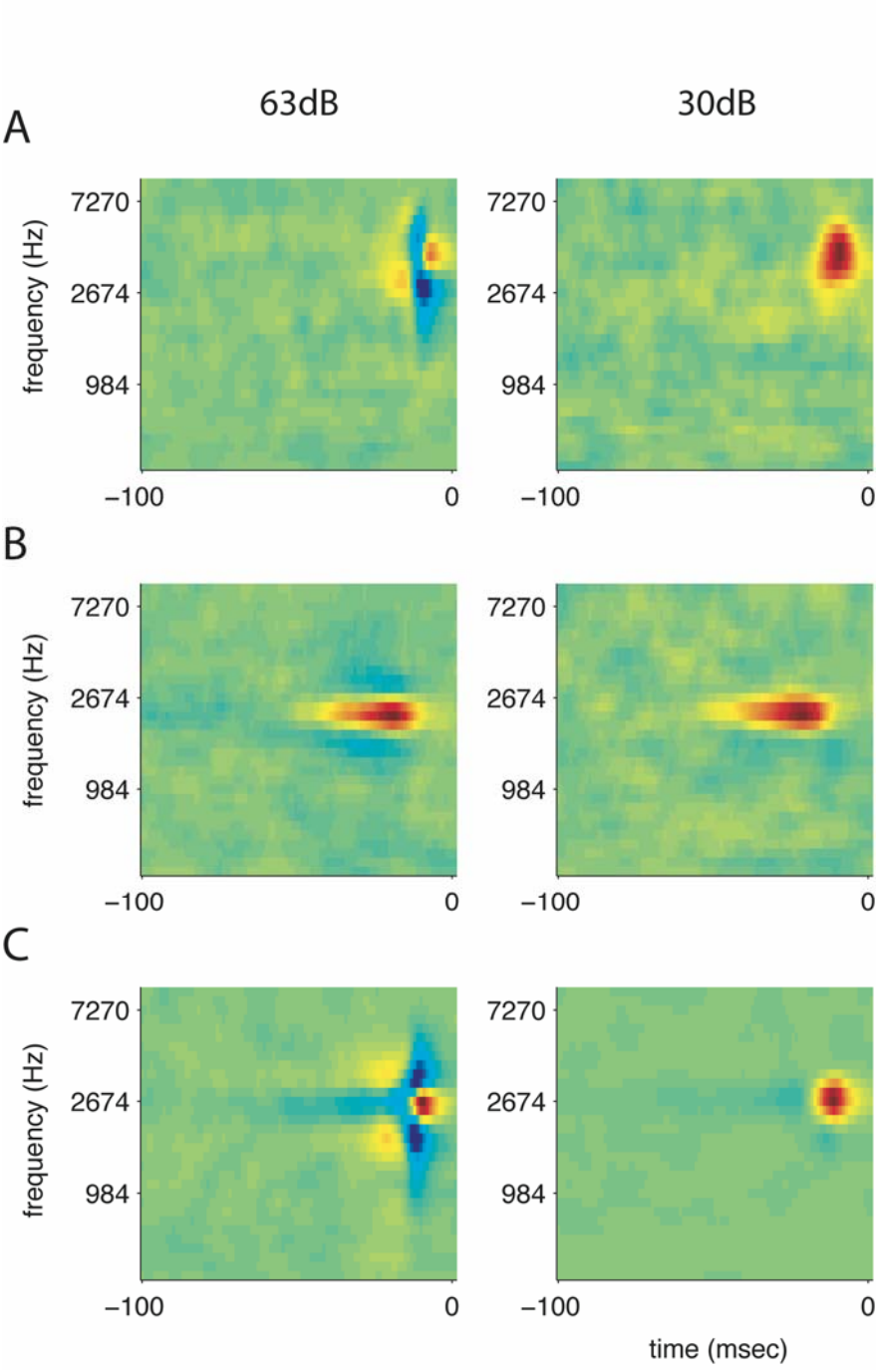
Neuron, volume 58

Supplemental Data

Organizing Principles of Spectro-Temporal Encoding in the Avian Primary Auditory Area Field L

Katherine I. Nagel and Allison J. Doupe

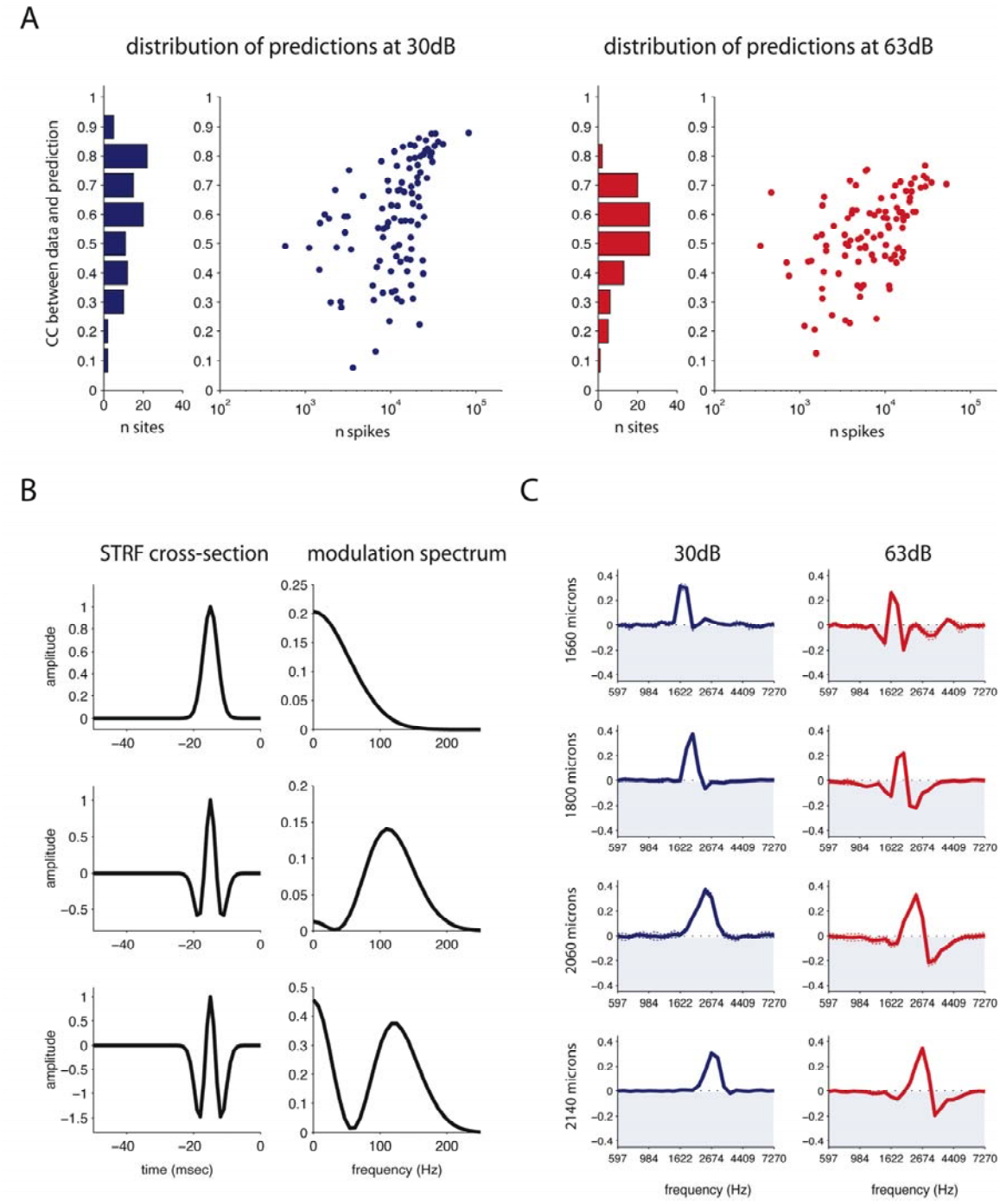
Supplementary Figure 1: Raw STAs for 3 example STRFs



Supplementary Figure 1

Raw (un-decorrelated) spike-triggered averages (STAs) for the example cells shown in figure 2A-C. At 63dB, the differences in tuning between a temporal (A), spectral (B), and spectro-temporal (C) cell are evident without decorrelation. At 30dB all three cells have much simpler STAs dominated by a single large positive peak.

Supplementary Figure 2



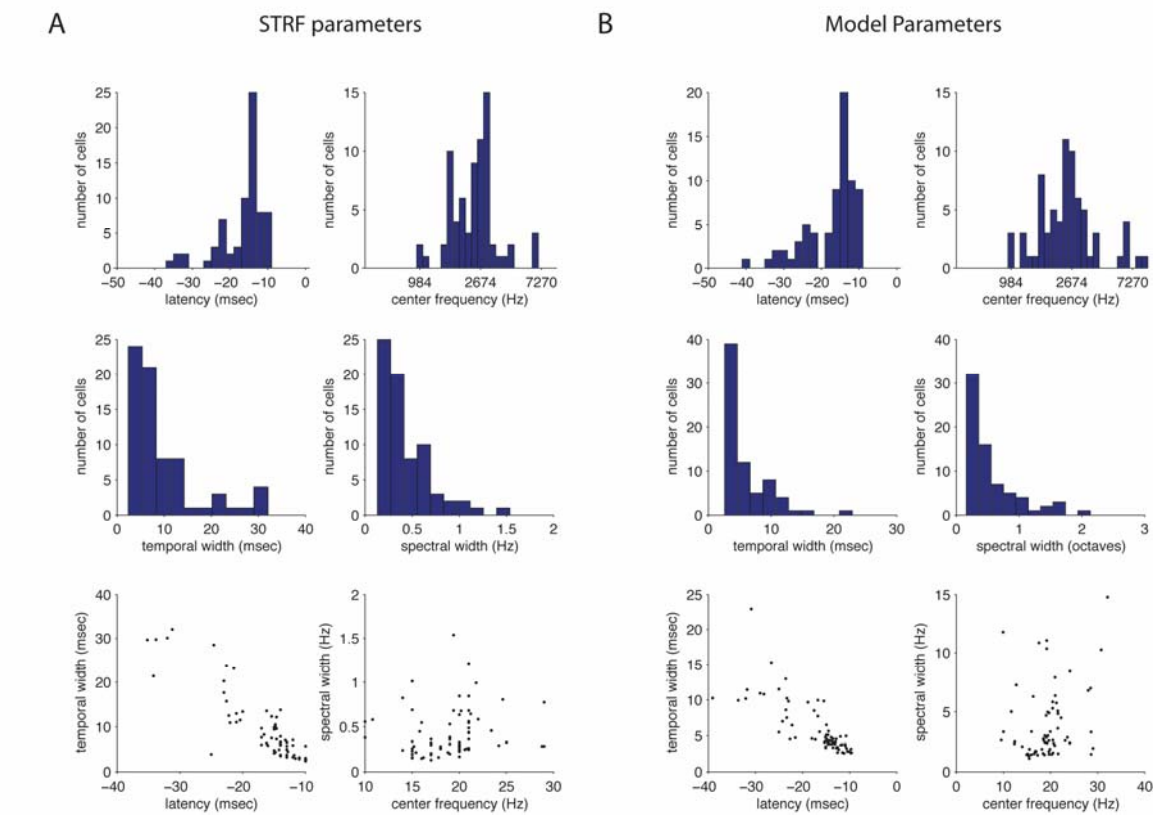
Supplementary Figure 2

A) The distribution of correlation coefficients between predicted and actual PSTHs for stimuli played at 30dB (left) and 63dB (right). PSTHs were obtained from 13 to 200 repeats of an identical 5 second stimulus segment by smoothing the average spike count per millisecond with a 8 msec-wide hanning window. Predictions were made using STRFs and nonlinearities derived from a non-overlapping set of data as described in the Methods. The first 500 msec of both the predicted and actual PSTH were omitted to reduce the effects of slow adaptation on the comparison. Prediction quality generally increased with the number of spikes collected, but plateaued at lower values for the 63dB stimulus condition. Better predictions may arise from cells with high firing rates for several (not mutually exclusive) reasons: because more data allowed better estimation of these STRFs; because cells with high firing rates tend to be sensitive to higher temporal frequencies; or because these cells are often located in a lower level of the processing hierarchy. The mean correlation coefficient was 0.57+/- 0.19 (standard deviation) for the 30dB stimulus condition, and 0.52 +/- 0.14 for the 63dB stimulus condition.

B) Different filter shapes give rise to different modulation spectra. The filter shapes shown represent either temporal or spectral-cross sections through the STRF, and the power spectra represent corresponding cross-sections through the two-dimensional modulation spectrum. If a filter has a single peak (top row), its power spectrum will be low-pass, meaning that the cell will respond to all modulations in that domain up to some cutoff determined by the width of the STRF peak. A filter with a single negative peak will give rise to a similar power spectrum. If the filter has balanced positive and negative regions (second row), its power spectrum will be band-pass, meaning that it will only respond to modulation frequencies over a narrow range. Finally, a filter with very profound negative sidebands (3rd row), whose negative regions are larger than its positive regions, will have a power spectrum with two peaks. The alternating positive and negative regions will give rise to a high frequency band-pass peak, while the overall negative sum will give the power spectrum a peak at low frequencies. This property of power spectra complicates our analysis of some strongly spectrally-tuned cells, such as the example in figure 2C, whose STRFs show clear sensitivity to spectral modulations, but whose spectral power spectra peak at 0 cycles per octave.

C) Spectral cross-sections through several spectro-temporal STRFs illustrating their similar structure. Most such cells had a single positive peak at 30dB, while at 63dB they showed strong negative sidebands. These sidebands were generally stronger and narrower on the high frequency side, and shallower and broader on the low-frequency side. The STRFs shown here were collected at the depths indicated, in a single bird, and suggest a systematic progression of preferred frequencies.

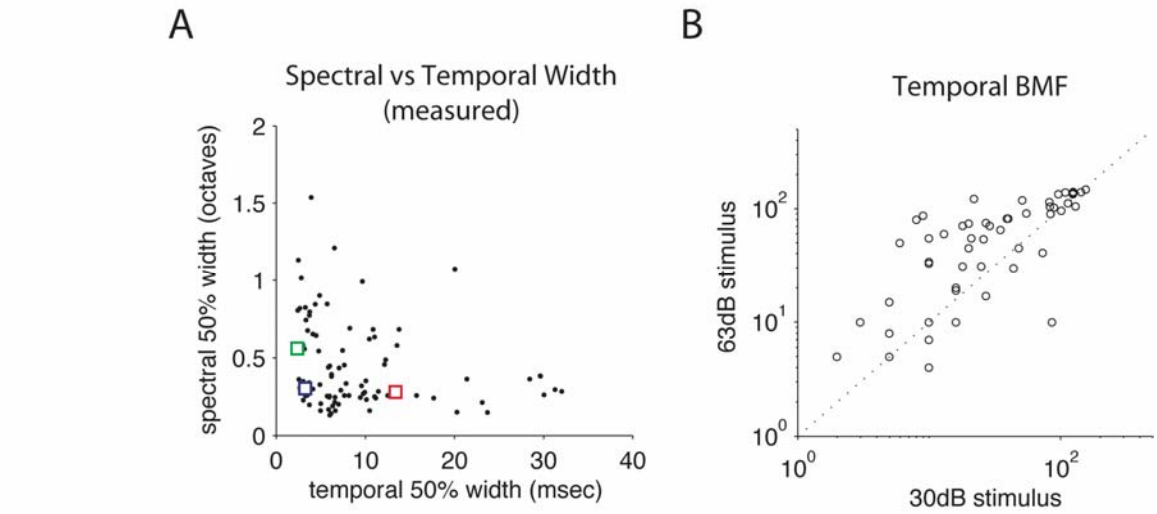
Supplementary Figure 3



Supplementary Figure 3

Distribution of STRF parameters obtained by measuring STRFs directly (A) and by fitting a bi-variate Mexican-hat model to each STRF (B). The measured latency and center frequency of the STRF were defined to be its maximum in time and frequency. The measured temporal half-width, and bandwidth were obtained from cross-sections through the peak. Model parameters were obtained by fitting as described in detail in the Methods. In general, the shape of the two distributions showed a good agreement. Latencies showed a skewed distribution with more cells having shorter latencies. Center frequencies showed an approximately normal distribution centered around 2.5 kHz. Temporal and spectral widths had more exponential distributions, with more cells having short widths. Latency and temporal width were negatively correlated, while center frequency and spectral width (in octaves) were not. Correlation coefficients between measured and fitted parameters were 0.97 for latency, 0.86 for center frequency, 0.89 for temporal widths, and 0.57 for spectral widths).

supplementary figure 4

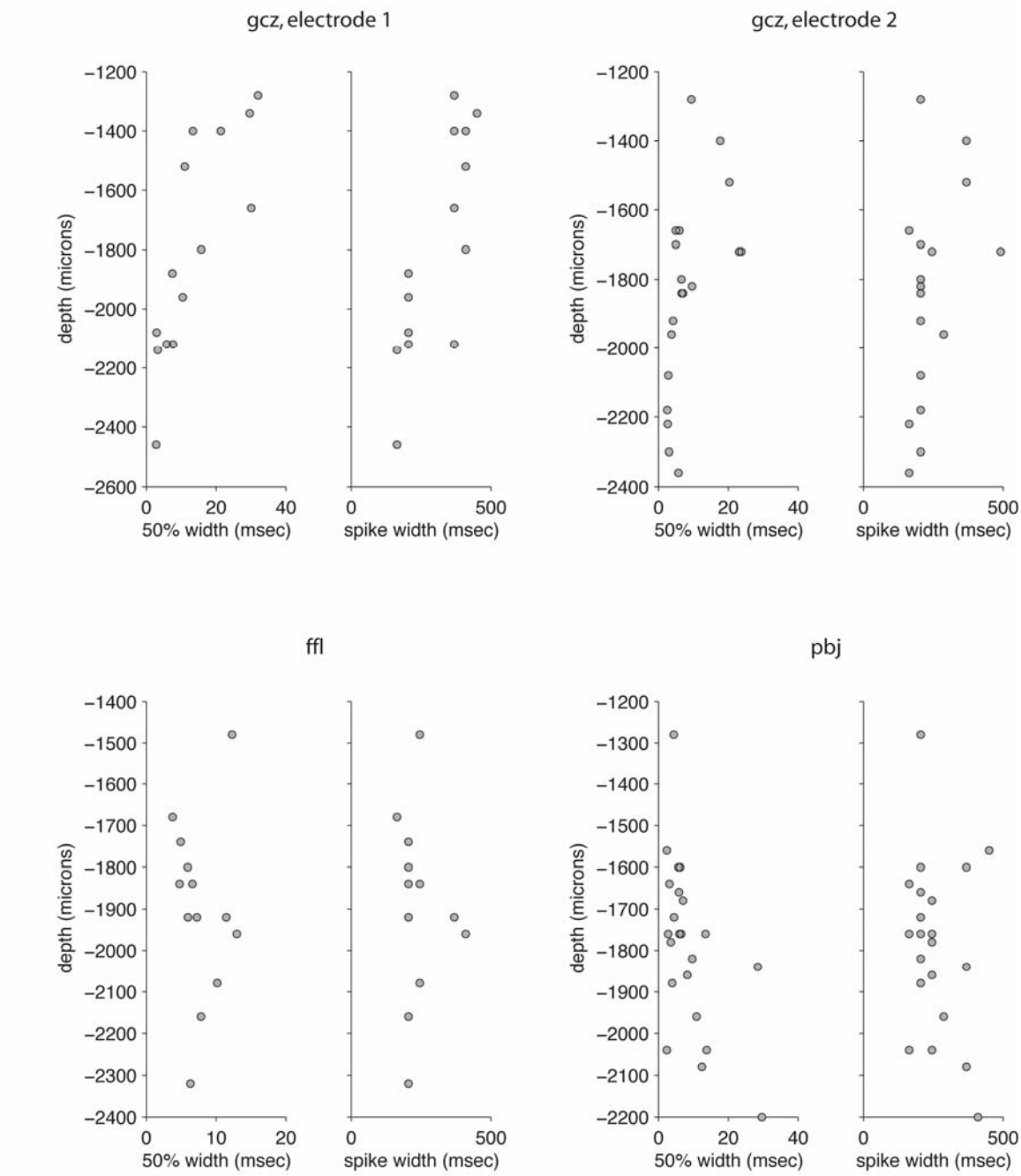


Supplementary Figure 4

A) Measured spectral versus temporal width for all significant STRFs (n=71) obtained at 63dB. Width parameters were obtained by taking temporal and spectral cross-sections through the peak of the STRF and measuring the interval over which the cross section was greater than half the maximum value. Example cells from figure 2 are indicated by green (temporal, panel 2A), red (spectral, panel 2B), and blue (spectro-temporal, panel 2C) squares.

B) Temporal best modulation frequency of STRFs at obtained at 63dB versus those obtained at 30dB. Temporal best modulation frequencies were obtained by summing the two sides of the modulation spectrum for each STRF, averaging across spectral frequencies, and finding the maximum of the resulting temporal power spectrum. Most points lie above the diagonal indicating that most cells become sensitive to faster temporal modulation frequencies when the stimulus intensity is higher.

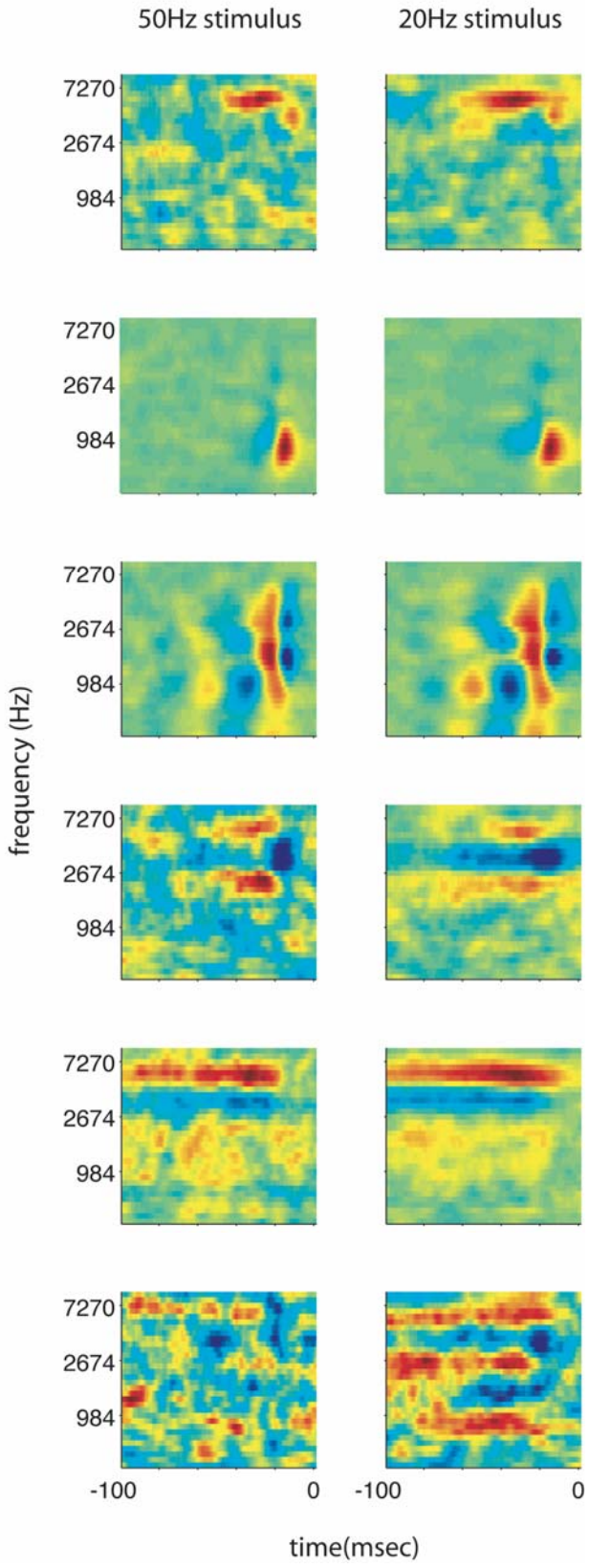
Supplementary Figure 5



Supplementary Figure 5

The anatomical distribution of STRF and spike properties from chronic birds supports our findings in restrained sedated birds (figures 7 and 8). This figure shows temporal half-widths of STRFs and widths of spikes as a function of recording depth for four chronic electrode penetrations. Fast STRFs with short temporal half-widths and narrow spikes are mostly found in a restricted region of each penetration. The strong correlation between a STRF's temporal width and spike width is also clear.

Supplementary Figure 6



Supplementary Figure 6

Spike-triggered averages recorded from the same site under two stimulus conditions, one with faster amplitude modulations (50Hz, left) and one with slower amplitude modulations (20 Hz, right). In general the STRFs obtained under the two conditions were very similar, although the STRFs obtained with the slower, 20Hz stimulus are smoother and slightly broader in both frequency and time. Compared to the striking effects of stimulus intensity on STRF shape (supplementary figure 1), the effects of stimulus frequency are very mild. Occasionally, as at the last site shown, sensitivity to very slow stimulus features was evident with the 20Hz stimulus but not with the 50 Hz stimulus.

Supplementary Analysis

Because the shape of an extracellular spike depends on the distance between the recording electrode and the cell (Gold et al, 2006), we worried that the relationship between spike width and temporal tuning could be due to differences in recording conditions. There was a weak correlation between spike width and spike amplitude in our data (cc = 0.26, p = 0.03), arising from a few units with large narrow or small wide spikes. To verify that the relationship between spike width and STRF width was not due to recording distance, we excluded cells with the largest and smallest amplitudes, below 200μV or above 600μV, leaving 60 of 71 cells in our sample. Spike and STRF width were still strongly correlated in this sub-population (cc = 0.74, p = 9.2-12), but spike width and spike amplitude were not (cc = 0.07, p =0.55). No significant relationship between spike amplitude and STRF width remained within this restricted population (correlation coefficient = 0.04, p = 0.79).

Supplementary Methods

Stimulus

For a subset of the acute mapping experiments we used a more slowly varying set of log envelopes to modulate the narrowband signals described in Figure 1 and Methods. The power spectrum of these log envelopes was given by

$$P(f) = \exp(-f/20\text{Hz})$$

STRFs estimated with the slower stimulus were generally smoother and less noisy but otherwise did not differ significantly from those estimated with the faster stimulus. Examples of STRFs from the same site estimated with both stimuli are shown in figure S6. For each mapping experiment we used only a single stimulus type.

STRF Estimation and Decorrelation

Our analysis attempted to relate the firing of field L neurons to a linear filter of the stimulus spectrogram. If this spectrogram were uncorrelated in both frequency and time, we could calculate this filter, or spectro-temporal receptive field (STRF), by simply cross-correlating the spike train with each row of the spectrogram, yielding the spike-triggered average, or STA. Because the spectrogram of our stimulus contains local correlations, this STA will instead be equal to the true neural filter, convolved with the autocorrelation of the spectrogram. The influence of these stimulus correlations can be removed from the STA by inverting the autocorrelation matrix, or "decorrelating" the STRF. Preparing our stimulus for analysis thus required two steps: first calculating its spectrogram, and then calculating the autocorrelation matrix of the spectrogram envelopes.

To calculate the spectrogram $S(cf,t)$, we first filtered the stimulus through the same bank of 32 log-Gaussian filters used to create the narrowband noise signals. We then extracted the envelope of each narrowband signal by passing its absolute value through the exponential filter used to generate the stimulus envelopes. We down-sampled the resulting low-pass signal to 1kHz to speed analysis, took its logarithm, and normalized it to have zero mean and unit variance. Because of the overlap in our frequency bands, the log envelopes of several of the original noise bands contributed to each of these extracted log envelopes and the extracted envelopes $S(cf,t)$ were not exactly equal to the envelopes used to create the stimulus.

The autocorrelation matrix of the spectrogram consisted of 128 complex matrices $A_\omega(cf, cf)$, representing the correlations between the log envelopes $S(cf,t)$ at each of 128 temporal frequencies ω. The diagonals of each matrix $A_\omega(cf, cf)$ represent the power in each log envelope at the temporal frequency ω, while the off-diagonal terms represent the correlation between log envelopes at that temporal frequency. To calculate these complex matrices, we broke $S(cf,t)$ into 256 sample-long chunks $S(cf,t:t+255)$, applied a hanning window to each chunk, then took its Fourier transform, yielding a complex matrix $\tilde{S}(cf,\omega)$. The complex matrix $A_\omega(cf, cf)$ was then equal to the average cross-product across chunks:

$$\left\langle \tilde{S}(cf,\omega)' \times \tilde{S}(cf,\omega) \right\rangle$$

To estimate each STRF, we first calculated the raw spike-triggered average $C_n(cf,\tau)$ by cross-correlating the spike train ρ(t) with each log envelope $S(cf,t)$ in the spectrogram:

$$C(cf,\tau) = \frac{1}{n}\sum_t \rho(t) S(cf, t-\tau)$$

where n is the total number of spikes. We obtained spike-triggered averages over an interval from 100msec before the spike to 100msec after. The portion of this average that followed the spike was used as an estimate of the noise in our averaging process.

To decorrelate the spike-triggered average, we first took its Fourier transform, yielding a complex matrix $\tilde{C}(cf,\omega)$. Each column of this matrix $\tilde{C}(cf,\omega)$ corresponding to a particular temporal frequency ω, was then decorrelated by applying the pseudo-inverse of the autocorrelation matrix for that temporal frequency $A_\omega(cf, cf)$. To apply the pseudo-inverse, we first took the singular value decomposition of the matrix $A_\omega(cf, cf)$:

$$U\Sigma V' = A_\omega(cf,cf)$$

where *U* and *V* are square matrices and Σ is a matrix of singular values that has non-zero values only along the diagonal. We then calculated a matrix Σ_{inv}, by taking the inverse of the values of Σ along the diagonal:

$$\Sigma_{inv}(s,s) = \frac{1}{\Sigma(s,s)}$$

We used Σ_{inv},with *U* and *V* to decorrelate the spike-triggered average:

$$\tilde{D}(cf,\omega) = V\Sigma_{inv}U'\tilde{C}(cf,\omega)$$

The inverse Fourier transform of this "decorrelated" matrix $\tilde{D}(cf,\omega)$ is equal to the decorrelated spike-triggered average $D(cf,\tau)$. This method is very similar to the one used by Theunissen et al. (2000).

A limit of this decorrelation technique is that it tends to boost the power in high temporal frequencies that are poorly sampled by the stimulus, leading to very noisy estimates of the STRF. This is because the inversion increases the magnitude of the smallest singular values. To limit this noise we systematically reduced the amplitudes of the largest inverse singular values. We first obtained the singular values of all the autocorrelation matrices and sorted them in descending order (largest to smallest). We then found the largest singular value of the autocorrelation matrix at a cutoff frequency $A_{\omega cut}$, whose choice is described below. We call this cutoff singular value s_{cut}. Singular values greater than this cutoff value were inverted as described above. Singular values below this cutoff were reduced in amplitude according to their distance from the cutoff value

$$\Sigma_{inv}(s,s) = \frac{1}{\Sigma(s,s)} \times \exp(-k/250)$$

where k is the number of singular values between the current value and the cutoff. By smoothly reducing the amplitude of large inverse singular values, this method reduces "ringing" in the STRF that can arise from using a hard cutoff on the decorrelation.

We tried several methods for choosing an appropriate cutoff frequency. The first was to find the frequency at which the power in the spike-triggered average fell below 3 times the mean power in the acausal part of the the

spike-triggered average. This "noise cutoff" yielded reasonable results for some cells, but unreasonably high cutoffs (150-250 Hz) for others. STRFs decorrelated using a cutoff frequency greater than 150Hz were uniformly noisy and impossible to interpret. We therefore tried an alternate method, finding a cutoff frequency that contained 99% of the power in the raw spike-triggered average. This "99% cutoff" yielded much better results for STAs with excessively high noise cutoffs, but was too high for those with low noise cutoffs. We therefore chose one of these two cutoffs for each STRF depending on whether they yielded a significant STRF. If the 99% cutoff produced a STRF with at least 20 points above the maximum absolute value of the acausal STRF, we used that cutoff (n=50/74). Otherwise we used the lower noise cutoff (n=24/74). We observed no significant differences between the groups of cells decorrelated with one cutoff or the other and results for the two groups are pooled. Un-decorrelated spike triggered averages for the example cells in figure 2A-C are shown in supplementary figure 1. Both differences in tuning between cells, and differences in the same cell's STRF with intensity, are evident in these raw averages.

Both the causal and acausal parts of the STA were decorrelated together. STRFs were considered significant if the causal part of the STRF contained at least 20 points that were greater than the maximum absolute value in the acausal part of the STRF. 74 of 81 single units produced significant STRFs in at least one condition: 71 using the 63dB stimulus, 68 using the 30dB stimulus. To obtain error bars on our STRF estimates we divided our spikes randomly into 5 groups. Separate STAs and decorrelated STRFs were calculated from each group of spikes although a uniform cutoff was used for all decorrelations. Error bars shown in the figures represent the standard deviation of the jackknife estimate (Sen et al., 2001)

$$std = \sqrt{\frac{n-1}{n}\sum_{j}(x_j - < x_j >)^2}$$

where n is the number of jackknife estimates, and x_j is the jth estimate of the STRF.

Bi-variate Mexican hat model fitting, and a model with orientation

To fit the bivariate Mexican hat model (see Methods) to our data, we used a constrained least-squares fitting algorithm based on the Newton method (lsqcurvefit.m, Matlab). To choose starting values for each of the

parameters, we first created a matrix of initial parameter values. For A, x, and y, there were two initial values, equal to the magnitude, and t- and f-coordinates, of the maximum and minimum point on each STRF. Initial values of α and β were 0, 0.25, 0.5, 1, 2, and 4. The parameters σ and γ were fit by fitting their inverses $1/\sigma$ and $1/\gamma$, to avoid the possibility of having a zero in the denominator of the function. Initial values for these two inverse parameters were 0.1 to 0.8 in steps of 0.1. We calculated the mean squared error between the actual STRF and the model produced by each of the 18432 combinations of initial parameter values. The combination of parameters that produced the lowest error was chosen as the starting point for the algorithm. The algorithm was constrained to find values of x and y that fell within the temporal and spectral dimensions of our STRF. The two width parameters σ and γ were constrained to be positive, and the two depth parameters α and β were constrained to be greater than or equal to zero. Confidence intervals for parameter estimates were made using the function nlparci.m in Matlab.

In addition, to estimate the orientation of each STRF, we fit each STRF recorded at 63dB to a version of the model containing an orientation term θ:

$$t' = (t - x)\cos\theta + (f - y)\sin\theta$$
$$f' = -(t - x)\sin\theta + (f - y)\cos\theta$$

$$G(t, f) = Ae^{-\sigma^2 t'^2 - \gamma^2 f'^2}(1 - \alpha^2\sigma^2 t'^2)(1 - \beta^2\gamma^2 f'^2)$$

The parameter θ was constrained to fall between –45 and +45° to avoid ambiguity between spectral and temporal parameters of the model. To fit this model, we initialized the first seven parameters to the values we found by fitting the simpler model, and set θ to zero. All parameters were allowed to vary during the second fitting.