Analyzing Variability in Neural Responses to Complex Natural Sounds in the Awake Songbird.

Authors and affiliation

Gilberto David Graña¹,²
Cyrus P. Billimoria¹,²
Kamal Sen¹,²

¹ Hearing Research Center & Center for Biodynamics
² Department of Biomedical Engineering, Boston University

Running head
Auditory Response Variability in Awake Songbirds

Contact Information
Kamal Sen
Hearing Research Center, Department of Biomedical Engineering.
Center for Biodynamics, Program in Mathematical and Computational Neuroscience.
Boston University.
44 Cummington Street
Boston, MA 02215
kamalsen@bu.edu

Abstract

Studies of auditory processing in awake, behaving songbirds allow for the possibility of new classes of experiments, including those involving attention and plasticity. Detecting and determining the significance of plasticity, however, requires assessing the intrinsic variability in neural responses. Effects such as rapid plasticity have been investigated in the auditory system through the use of the spectrotemporal receptive field (STRF), a characterization of the properties of sounds to which a neuron best responds. Here we investigated neural response variability in awake recordings obtained from zebra finch field L, the analog of primary auditory cortex. To quantify the level of variability in the neural recordings, we used three similarity measures; an STRF-based metric, a spike train correlation-based metric, and a spike train discrimination-based metric. We then
extracted a number of parameters from these measures, quantifying how they fluctuated over time. Our results indicate that 1) awake responses are quite stable over time; 2) the different measures of response are complimentary; specifically, the spike train based measures yield new information complementary to the STRF; and 3) different STRF parameters show distinct levels of variability. These results provide critical constraints for the design of robust decoding strategies and novel experiments on attention and plasticity in the awake songbird.

**Keywords:** Auditory cortex, speech, birdsong, field L
INTRODUCTION

An animal’s accurate responses to natural stimuli are often essential to its survival. The study of how these stimuli are encoded and processed at the neural level should provide insight into the fundamental problem of neural coding. Of specific interest are natural sounds, which display intricate time-varying structure over multiple time-scales (Attias and Schreiner 1998; Escabí et al. 2003; Lewicki 2002; Nelken et al. 1999; Singh and Theunissen 2003) that provide important cues for sound recognition. Previous studies have suggested that the auditory cortex plays an important role in the processing of natural auditory scenes and species-specific vocalizations (Fitch et al. 1997; Nelken 2004; Rauschecker 1998; Wang 2000).

Songbirds provide a model system that offers unique advantages for studying the processing of complex sounds (Doupe and Kuhl 1999). As with other animal species, songbirds communicate using a set of acoustically complex vocalizations which transmit behaviorally relevant information important for intra-species interactions such as descriptions of the identity and location of the sender (Konishi 1985; Theunissen et al. 2000; Wang 2000). Field L, the avian analog of the auditory cortex, is selective to conspecific vocalizations and is thought to play an important role in the processing and discrimination of these sounds (Grace et al. 2003; Sen et al. 2001; Theunissen et al. 2004; Theunissen et al. 2000).

How does one assess the variability of neural recordings? The awake system allows for the investigation of both attentional effects and rapid, online plasticity, but is subjected to
neural response variability over time. Thus, answering this question is vital to understanding and validating the results of long-term neurophysiological experiments, including short-term plasticity effects caused by behavioral training (Eggermont 2006; Fritz et al. 2003; Spierer et al. 2007; Weinberger et al. 2006). An accurate assessment of the variability of neural responses also allows for the development of robust decoding strategies. To answer this question, we chose to investigate the spectro-temporal receptive field (STRF), a widely used characterization of the spectral and temporal features of sounds to which neurons best respond (Escabí and Schreiner 2002; Sen et al. 2001; Theunissen et al. 2000). An elegant set of studies have employed the STRF to investigate rapid plasticity in neural responses, revealing dramatic task-dependent plasticity in primary auditory cortex (Elhilali et al. 2007; Fritz et al. 2005; Fritz et al. 2003). Here we obtain STRFs in awake restrained zebra finch field L using complex natural sounds, specifically conspecific (same species) vocalizations, and investigate the intrinsic variability of STRFs over time, up to several hours.

We also employ two other response measures: a spike train correlation-based metric and a spike train discrimination metric (Billimoria et al. 2008; Larson et al. 2008; Narayan et al. 2006; Wang et al. 2007) to quantify variability. We address several questions: 1) How stable is the STRF over time? 2) Do other quantitative measures of response yield complementary information on response variability? 3) Do different parameters of the STRF show identical levels of variability?
METHODS

Surgical procedures

We recorded from the field L region of awake adult male zebra finches (*Taeniopygia guttata*). All procedures were in strict accordance with the National Institutes of Health guidelines as approved by the Boston University Charles River Campus Institutional Animal Care and Use Committee. Two days prior to the electrophysiological recording, the bird was anesthetized (0.1 – 4% isoflurane in 0.5-2.5 l/min O₂) for a preparatory surgical procedure to implant a manually-operated microdrive and fix a head support pin. A reference point for electrode penetrations was marked with ink 1.5 mm lateral and 1.2 mm anterior to the bifurcation point of the midsagittal sinus. A lightweight microdrive containing two extracellular tungsten electrodes (impedance: 2-4 Ω, FHC, Bowdinium, ME) was positioned above the marked dot. The microdrive was a modified version of a previous device used to record from awake zebra finches (Hessler and Doupe 1999). The skull and dura beneath the dot were removed, and the implant was positioned such that the electrodes just entered the brain. The implant was secured to the skull in this configuration with epoxy. A reference ground electrode was inserted into the brain on the opposite hemisphere from the location of the implant. Finally, a steel support pin was glued to the skull above the midsagittal sinus. The bird was allowed to recover for 2 days before performing the experiments.

Stimuli

The stimulus ensemble played to the subjects consisted of 20 undirected, conspecific zebra finch songs recorded in a sound-attenuated chamber (Acoustic Systems, Austin,
The songs were sampled at 32 kHz, band-pass filtered to retain frequencies between 250 Hz and 8 kHz, and stored in data files for playback (Sen et al. 2001).

**Electrophysiology**

On the day of the experiment, the bird was restrained in a small cloth jacket to limit movement and reduce motion artifacts; the bird was then positioned in a stereotactic assembly and placed into a double-walled sound-attenuated chamber (Industrial Acoustics, Bronx, NY), facing a loudspeaker that was used for stimulus presentation. The speaker was located 20 cm away from the beak, and the bird was elevated to be at the same height as the center of the speaker cone. The bird's head position was fixed by attaching the steel pin to a frame located on the stereotactic assembly; this served to further reduce the incidents of motion artifacts in the recordings. Single- and multiunit complexes in the same adult zebra finch were probed by manually advancing the two tungsten electrodes via the microdrive in approximately 150 µm intervals and playing a conspecific song that was not part of the test stimulus ensemble. The bird was given a 30-min break after each recording session not lasting more than three hours and released from the cloth jacket into its cage. Data from a particular site were obtained within the same recording session; occasionally, some sites were probed and responses recorded over two or more successive sessions. Different sites were sampled over several days to weeks. After the experiment was concluded, the bird was euthanized with an overdose of isoflurane, and the brain was preserved in 3.7% formalin fixative for histology.
**Histology**

Prior to sectioning, the brains were stored in 30% sucrose buffer overnight. Parasagittal 50 μm sections of the brain were prepared using a cryo-microtome and stained with cresyl violet (Nissl stain). Electrode placement was verified by comparing electrode tracks and electrolytic lesions to histological markers that define the boundaries of field L (Fortune and Margoliash 1992). Sites were classified as field L sites based on a combination of histology, medial-lateral coordinates and depth of the recording site.

**Neural Data**

Data were analyzed for STRF calculation only for sites that exhibited an average firing rate that was significantly different ($p < 0.001$, paired t-test) from the average spontaneous firing rate for at least one song stimulus. At a number of these sites, we obtained multiple recording blocks, defined as the interleaved presentation of 10 trials each of 20 conspecific zebra finch songs. Each recording block was presented at least 20 minutes after the start of the previous recording block for a particular site. Recording sessions were terminated early if the subject became restless or if three hours had expired. In this manner, neural data were obtained from 28 sites in 9 birds. Trials that demonstrated motion artifacts were removed from the data sets, though these comprised a small number of the total number of trials (approximately 0.5%).

Spike sorting and data analysis were performed using custom software written in Matlab (Mathworks, Natick, MA). Spike event times were obtained from the spike waveforms using a visually determined threshold and a window discriminator; the same threshold...
was used for spike event times across multiple recordings from a particular electrode depth and bird. Classification of sites into multiunits followed the scheme used in Sen et al. 2001. Multiunits consisted of spike waveforms that could be easily distinguished from the background noise but not from each other and contained small clusters of neurons.

**STRF Calculation**

A detailed description of the calculation of STRFs from natural sounds can be found in Theunissen et al. (2000). A STRF can be defined as the optimal linear filter that transforms a particular representation of a time-varying stimulus into a prediction of a neuron’s firing rate. Calculations were performed using the STRFPACK software suite (Zhang et al. 2006), which uses an invertible spectrographic representation of the sound stimuli and finds the optimal linear filter that, when convolved with the spectrographic representations of the stimuli, best matches the predicted firing of the recording site. This method corrects for the spectral and temporal correlations present in the sounds by performing a calculation that normalizes the song-spike train correlation by the autocorrelation of the stimuli. The resultant STRF is presented with limited-resolution in the spectral and temporal domains both as a result of the calculations performed and limited by the resolution of the inputs; in particular, the spectral resolution is 250 Hz (the bandwidth of the filter banks used to decompose the song spectrogram) and the temporal bin width is 1 ms (as the sampling rate of the spike trains is 1 kHz).

To determine the significant regions in the STRFs obtained, a jack-knife resampling method was used (Sen et al. 2001); STRFs were calculated for multiple subsets of the
conspecific song ensemble, obtained by deleting one song at a time from the entire ensemble. The variance for each spectral-temporal bin in the STRF estimate was calculated from this set of STRFs; the standard error was obtained by taking the square root of the variance and was plotted on the STRF images.

The STRF was de-noised using singular-valued decomposition (SVD). Singular values were calculated for the causal (0 to 100 ms) and acausal (-150 to -50 ms) regions of the STRF. The singular values from the causal region greater than the maximum singular value from the acausal region were used to reconstruct a de-noised STRF. We also calculated the signal-to-noise ratio (SNR) of this de-noised STRF by dividing the power in this causal region by the power in the acausal region.

**STRF-based Similarity Index**

We quantified the level of variability between two STRFs calculated from different recording blocks at the same site using a similarity index ($SI$) (DeAngelis et al. 1999; Escabí and Schreiner 2002). First, the STRFs in question were treated as vectors by reading the values from each matrix down its columns. These vectorized STRFs were then used to calculate the similarity index:

$$SI = \frac{\langle STRF_A, STRF_B \rangle}{|STRF_A| |STRF_B|}$$

where $\langle, \rangle$ is the vector inner product and $|$ $|$ is the vector norm operator. Values of $SI$ close to 1 indicate strong similarity between the two STRFs. The $SI$ is calculated between
the first STRF and any successive STRFs obtained from the recording site under investigation.

**Correlation-based Similarity:** $R_{corr}$

To more robustly assess the variability in subsequent recordings, we employed two more computational measures derived directly from the neural recordings as opposed to the STRFs. The first of these, called $R_{corr}$, is based on a previously described correlation-based measure of spike similarity (Schreiber et al. 2003). The similarity between two spike trains $\vec{r}_i$ and $\vec{r}_j$ is calculated as follows:

$$ R_{corr} = \frac{\vec{s}_i \cdot \vec{s}_j}{||\vec{s}_i|| ||\vec{s}_j||} $$

where $\vec{s}_i$ and $\vec{s}_j$ were obtained by filtering $\vec{r}_i$ and $\vec{r}_j$, respectively, using a Gaussian filter with mean 0 and standard deviation 10 ms. An $R_{corr}$ value close to 1 indicates similarity between the spike trains, whereas a value close to 0 indicates dissimilarity. Averaging the values of $R_{corr}$ over the different pairs of spike trains obtained in one recording yields a measure of the reliability of the unit in responding to different stimuli.

We used the $R_{corr}$ measure in two ways. First, as described in the previous section, we calculated $R_{corr}$ as the reliability of responses obtained within recording blocks. Secondly, we adapted $R_{corr}$ to quantify the similarity between sets of neural responses obtained across recording blocks at the same site. To adapt the $R_{corr}$ measure for this series of calculations, $\vec{r}_i$ and $\vec{r}_j$ represent spike trains resulting from presentation of the same stimulus but obtained from different recording blocks at the same site. As for the STRF
similarity index, the calculation was also performed between spike trains in the first recording block and spike trains in each of the subsequent recording blocks. The resulting $R_{corr}$ values are then averaged over the trials and songs to obtain a measure of spike train similarity across the two recordings blocks.

**Performance-based Similarity**

The second similarity measure derived directly from the neural data is based on a previously described spike distance metric (SDM) (Narayan et al. 2006; van Rossum 2001). First, spike trains were filtered using a decaying exponential kernel with time constant $\tau$:

\[
f(t) = \sum_{i}^{M} H(t - t_i) e^{-\frac{(t-t_i)}{\tau}}
\]

where $t_i$ is the $i$th spike time, $M$ is the total number of spikes given a spike train length $T$, and $H(t)$ is the Heaviside step function. The spike distance is then computed between a pair of filtered spike trains $f$ and $g$:

\[
D^2(f, g) = \frac{1}{\tau} \int_0^\infty [f(t) - g(t)]^2 dt
\]

By varying $\tau$, discrimination could be measured over different time-scales of the neural response. At short time scales, the metric acts like a coincidence detector, with small differences in spike timing contributing to the distance, whereas at long time scales, the metric acts like a rate difference counter, where average firing rates contribute to the distance.
A classification scheme based on the SDM was then used to quantify the neural
discrimination of songs (Machens et al. 2003). Ten trials of spike trains were obtained
from the recording for each of the 20 songs aligned at their onsets. A template spike train
was chosen for each song, and remaining spike trains were assigned to the song with the
closest template based on the spike distance measure. This procedure was repeated 1000
times for different templates. Discrimination performance was quantified by computing
the percentage of correctly classified songs (% correct). The chance level for
classification is 5%, since a spike train could be assigned to 1 of 20 songs.

As with $R_{corr}$, a percentage correct was calculated in two ways for the various recordings
obtained from a particular neural unit. The first was as described above, selecting
template spike trains and assignment spike trains from the same recording block.
Secondly, the metric was adapted to act as another measure to quantify the similarity
between neural responses in blocks from the same recording site. For this set of
calculations, $f$ and $g$ are filtered spike trains from two different recording blocks; the first
recording block made at a particular site is the reference, $f$, and $g$ is the set of responses
from each subsequent recording block obtained at that site (e.g., distances are calculated
between spike trains from the first recording block and the second recording block, then
from the first recording block and the third recording block, and so on). The value of $\tau$
was chosen to reflect the optimal timescale observed in previous experiments (Narayan et
al. 2006; Wang et al. 2007); the standard deviation in the $R_{corr}$ calculations above was
chosen to coincide with this value. The parameters for calculating the distances were
fixed, with $\tau$ at 10 ms and a spike train length of 1000 ms. The template songs were then
chosen from the spike trains of the first recording block, and spike trains from the subsequent recording blocks are matched to this template. This procedure was repeated for 1000 different permutations of template songs, generating a percentage correct of correctly classified songs.

Analysis of STRF Parameters

We also analyzed STRF similarity by studying the changes in six parameters derived from the STRF as described previously (Sen et al. 2001). Briefly, these parameters are latency, the time of the peak amplitude in the STRF; the center frequency (CF), the frequency of the peak amplitude in the STRF; the quality factor (Q), a measure of the sharpness of this spectral peak; the excitatory-inhibitory ratio (EIR), which gives an estimate of the concentration of the excitatory and inhibitory energies in the STRF; the separability index (SI), a term derived from the singular value decomposition of the STRF, which quantifies the time-frequency separability of the filter; and the temporal integration window (TiW), which estimates the temporal characteristic of the neuron that processes the amplitude envelope of sound. One difference between the sets of parameters obtained is in choosing the temporal integration window over the best modulation frequency (BMF, the frequency showing the highest value in the power-spectral density obtained from a temporal slice obtained at the center frequency) which was used in Sen et al. We observed that, for individual recordings at many neural sites, the multiple jackknife estimates showed high variability in their BMF, and so we instead opted to use the temporal integration window, which gives similar information. Note that the time-frequency resolution chosen for the STRF, which was the same resolution used
in prior work on field L (Sen et al. 2001; Theunissen et al. 2000), affects the resolution of
the parameters extracted from the STRF. In particular, the minimum detectable change
for the CF is 250 Hz as a result of the filter banks chosen. This results in a 16.7% change,
for example, if the initial CF is 1500 Hz.
RESULTS

Similarity Index

We performed a total of 94 recordings from 28 sites in 9 zebra finches. For each of these recordings, STRFs were calculated and compared using the similarity index measure described above. Figure 1(A-D) shows four STRF plots obtained from one neural recording site. Areas in red indicate excitatory sub-regions, while areas in blue indicate inhibitory sub-regions. Dotted contours show one standard deviation, and solid contours show two standard deviations away from the mean STRF. The time stamp in the upper-right corner of each STRF plot indicates the time elapsed after the start of the first recording block obtained at the site. The four STRFs share qualitatively similar features, e.g., the time- and frequency of the main excitatory sub-region, with the shapes of the significant regions as depicted by the contours.

One of our goals was to use the information found in the STRF to derive a STRF similarity measure to analyze variability in the awake neural recordings. A majority of the recording sites showed high similarity over varying lengths of time as measured by the STRF similarity index. Figure 1(E) shows a graph of the similarity indices calculated between the first STRF and all subsequent STRFs for a particular recording site plotted against the time elapsed after the start of the first recording block obtained at each site with more similar STRFs having a similarity index closer to a value of one. The site depicted in panels (A) through (D) is outlined in red, and different sites are distinguished by a unique color/shape pair. On the right of the graph, a histogram shows the similarity index data binned in groups of 0.05, highlighting the number of STRFs for which there
was a high similarity index. The mean value of the similarity indices was 0.90 with a standard error of 0.01. We also calculated the percentage change in STRF similarity index values for sites where more than two recording blocks were obtained. The mean percentage change in STRF similarity index was -3.2%, with a standard error of 1.2%

Although a number of recording sites demonstrated qualitatively similar STRFs across recording blocks, there were a small number of sites where the STRF varied considerably across time. Supplemental Figure 1 shows an example of a variable STRF. The timestamps above each plot show the time of the recording relative to the start time of the first STRF. Text in the upper-left hand corner of each plot indicates the similarity index calculated between that STRF and the first STRF shown; these numbers are plotted with the upward-pointing, blue triangle symbols in Figure 1.

**Correlation-based Similarity: \( R_{\text{corr}} \)**

Response similarity was also quantified using the \( R_{\text{corr}} \) measure described above. Specifically, we calculated \( R_{\text{corr}} \) using pairs of neural recordings and plotted the results in Figure 2. Panels (A) through (D) show four sets of ten spike rasters each; these rasters correspond to the STRF plots in Figure 1. The time stamps above each set of rasters denote the time elapsed after the start of the first recording block at that site. Across both single-block trial recordings and the four recording blocks, the spike rasters show reproducible firing in response to the stimulus across a span of more than three hours.
We also quantified the similarity in these spike trains by calculating the $R_{\text{corr}}$ variability measure across recording blocks. Across the spike trains for all of the different recording sites, we observed low variability using the across-recording block $R_{\text{corr}}$ measure over time. Figure 2(E) shows a plot of the $R_{\text{corr}}$ measure calculated between the responses from the first recording block and the responses from each subsequent recording block at a particular site plotted against the time elapsed after the start of the first recording block for each site. A maximal value of one indicates perfect correlation. The example site depicted in panels (A) through (D) is outlined in red, and different sites are distinguished by a unique color/shape pair corresponding to the color/shape pair shown in Figure 1(E). On the right of the plot, a histogram shows the across-recording block $R_{\text{corr}}$ data binned in groups of 0.05. The histogram indicates a more spread out distribution of values as compared to the similarity indices. The mean value of $R_{\text{corr}}$ was 0.76 with a standard error of 0.01. We also calculated the percentage change in across-recording block $R_{\text{corr}}$ for sites where more than two recording blocks were obtained. The mean percentage change in across-recording block $R_{\text{corr}}$ was -1.7%, with a standard error of 0.8%.

To better understand the source of variability observed in the $R_{\text{corr}}$ values, we plotted across-recording block $R_{\text{corr}}$ against within-recording block $R_{\text{corr}}$ in Figure 2(F). For each site, we plotted all but the first within-recording site $R_{\text{corr}}$ value against the corresponding across-recording block $R_{\text{corr}}$ values as there is one less across-recording block value. A black line is plotted as the best-fit estimation between the data; the correlation coefficient was 0.98 ($p < 0.001$).
**Performance-based Similarity**

The third similarity measure used was based on the discrimination performance algorithm described above using the van Rossum metric. Many of the recording sites show low variability using the performance-based measure as indicated by similar values of the measure over time. Figure 3(A) shows a graph of the discrimination performance calculated between the responses from the first recording block and the responses from each subsequent recording block at a particular site plotted against the time elapsed after the start of the first recording block obtained at the same site. The site depicted in Figure 1(A-D) is outlined in red, and different sites are distinguished by a unique color/shape pair corresponding to the color/shape pair shown in Figure 1(E). On the right, a histogram shows the distribution of across-recording block discrimination values binned in groups of 5%. As with the $R_{corr}$ histogram, the values of the discrimination performance-based similarity values are more spread out than the similarity index values. The mean value of the discrimination performance was 64.87% with a standard error of 3.992%. As with STRF similarity index and $R_{corr}$, we calculated the percentage change in the performance-based similarity measure for sites where more than two recording blocks were obtained. The mean percentage change was 0.3%, with a standard error of 4.9%

We plotted across-recording block performance against within-recording block performance in Figure 3(B). As with the $R_{corr}$ calculations, for each site we obtained one more within-recording block measure than we did for the across-recording block calculation, and so we plotted all but the first within-recording block performance value against the corresponding across-recording block performance values. A black line is
plotted as the best-fit estimation between the data; the correlation coefficient was 0.95 ($p < 0.001$).

**Comparing the Similarity Measures**

We quantified the relationship among across-recording block $R_{corr}$, the performance-based similarity measure, and the STRF similarity index by plotting one against the other in Figure 4, showing that the three measures were highly correlated. Panel (A) shows the relationship between the performance-based similarity and $R_{corr}$; a least-squares linear fit yielded the line in black; the correlation coefficient was found to be 0.95 ($p < 0.001$). Panel (B) shows the comparison between across-recording block discrimination performance and the STRF similarity index; the correlation coefficient between these two measures was 0.69 ($p < 0.001$). Finally, panel (C) shows across-recording block $R_{corr}$ versus STRF similarity index; the correlation coefficient was 0.66 ($p < 0.001$). One recording site in particular – indicated by the red-filled circles – showed a marked difference between the two spike-variability measures, with high values for the STRF similarity index but lower values for the $R_{corr}$ measure and the performance-based measure. The STRFs for this recording site demonstrated qualitatively similar features across the recording blocks but showed a decrease in the absolute intensities of the peaks with time. This is further reflected in the spike rasters, which showed a decrease in the overall firing rate across recording blocks.

We also calculated the signal-to-noise ratio (SNR) for the STRF, comparing the power in the acausal region (defined as the STRF data between -150 and -50 ms) to that in the
causal region (defined as the STRF data between 0 and 100 ms) and compared the results to the three similarity measures. Supplemental Figure 2 shows the SNR plotted in dB for the second and subsequent STRFs obtained from a particular recording site plotted against the similarity indices calculated previously. The solid black line represents a significant correlation between the SNR and the similarity index ($r = 0.60, p < 0.001$).

We also compared the SNR to both the $R_{corr}$ and performance-based measures; both correlations were smaller and less significant than the one comparing the SNR and the STRF similarity index ($R_{corr}: r = 0.29, p = 0.018$; performance measure: $r = 0.35, p < 0.004$; data not shown).

**STRF Parameters**

We extracted a set of six parameters to characterize the STRF, providing an additional measure of similarity to compare multiple neural recordings from a particular site. The parameters are summarized in the plots of Figure 5. Each panel shows one of the six parameters plotted against the elapsed time since the first neural recording, with different units designated by different color/shape pairs. As before, the data corresponding to the example site featured in the figures are outlined in red. Percentage changes in these parameters, taken relative to the first recording for each site, are presented in Table 1. The plots demonstrate a wide range of values for the different parameters, indicative of the large variety of STRFs obtained in the study. Further, qualitative inspection of the parameters indicates that three of the six – Latency, excitatory-inhibitory ratio, and separability index – did not widely vary across recording times. The temporal integration window showed modest variability, while the center frequency and Q-factor showed
somewhat larger variations (however, see Methods for a description of how spectrotemporal resolution affects these parameters). No significant correlations were found between the percentage changes in the parameters and the time between recordings. We also investigated the relationship among the STRF parameters and the three similarity measures. Because there was one more value of each STRF parameter per recording site than resulting similarity values (which are computed pair-wise), all but the first value of each STRF parameter was paired to the corresponding similarity values for each recording site. The Pearson correlation coefficients for these pairings, as well as those resulting from comparisons between different STRF parameters, are summarized in Table 2.

**Changes in variability measures**

Figure 6 illustrated the percentage change in the values obtained for the different measures of variability. The three panels plot the changes in similarity index (A), across-recording block $R_{corr}$ (B), and the performance-based similarity measure (C); the percentage change was obtained by subtracting the second (and subsequent) variability measure value from the first and dividing by the value of the first measure, and is plotted against the corresponding elapsed time for the second (and subsequent) variability measures (this naturally disqualifies recording sites where only one time-point was obtained). A majority of the recording sites show a relatively low fractional change in the similarity value, with 50% of all units showing changes of less than 5% in STRF similarity index, 54% showing less than 5% change in across-recording block $R_{corr}$, and 50% showing less than 15% change in the performance-based measure. In each plot, a
unique symbol/color combination corresponds to a distinct recording site, matching up to the same symbol/color combination in previous figures. The dashed red lines correspond to the standard error of the mean.
DISCUSSION

Our analysis of the different variability measures indicated that awake responses show stability over multiple recording blocks. The three different measures – one derived from the STRF, a second from the $R_{corr}$ calculation, and a third from the distance metric calculation – are complementary to each other, and thus together give a quantitative measure of neuronal variability. We also discovered that not all STRF parameters showed similarity in the way they varied across recording blocks, as some were found to vary more than others, regardless of how the overall STRF behaved. By analyzing these similarity measures in future experiments, we can develop a method by which the overall variability of a recording site is quantified, giving us a sense of how viable plasticity- and attention-based experiments would be at that site. Furthermore, the results presented here can be used to develop decoding strategies that accommodate the natural variations seen in the neural responses.

This study compared three different measures for investigating neural response variability: one based on the neuron's STRF (Sen et al. 2001; Theunissen et al. 2000; Zhang et al. 2006), another based on a correlation-based spike similarity algorithm (Schreiber et al. 2003), and the third on a spike distance metric (Machens et al. 2003; van Rossum 2001). Analyses of these data involved comparing the results of the spike similarity and distance matrices when calculated within a particular recording block to these calculations taken across recording blocks at the same recording site. We also quantified the correlations between all pairs of variability measures to assess the contributions of neural response properties or specific firing patterns to neuronal variability. Further, by quantifying the
change in the individual variability measure values obtained from the different recording sites, we determined that neural recordings which demonstrated a particular measure of variability tended to maintain that level of variability as the time between neural recordings at that site increased.

Although auditory cortex has been a successful model for plasticity studies (Buonomano and Merzenich 1998; Weinberger 2007), relatively few studies have investigated rapid on-line cortical plasticity in an awake animal (Fritz et al. 2003). In particular, to our knowledge only one other study has quantified the stability and variability of cortical responses in the awake animal using the STRF (Elhilali et al. 2007). Our work serves to expand upon and supplement this study, providing additional measures that can serve to quantify plasticity effects.

**STRF-based Similarity**

The results of this study indicate that the similarity index (DeAngelis et al. 1999) groups a number of neural recording sites as relatively stable across recording blocks (Figure 1). This suggests that, at the cortical level, recording sites can be expected to react similarly across time if one takes into consideration the STRF alone. Though no hard criterion for establishing a “stable” unit was established here, another recently published paper (Elhilali et al. 2007) quantified “stable” versus “labile” STRFs using a tree-structured vector quantization (TSVQ) algorithm. This approach clusters STRFs based on their Euclidean distance from a so-called “standard” STRF; minimizing the cost function that incorporates different STRFs from this standard segregates experimentally-derived
STRFs into clusters. Units where repeated recording sessions yield STRFs in different clusters would be considered “labile”, while those grouped together would be considered “stable”.

Unlike the Elhilali study (2007), however, our variability measures do not use a pre-existing library of STRFs to assess the amount of similarity between STRFs obtained from a particular recording site. Instead, comparisons are made between the STRFs obtained from the recording sites themselves, so that notions of “stability” and “lability” are dependent on the properties of the recording site alone and not on a categorization to a family of STRFs. Further, our analysis is based not only on the STRF features, but is also based upon measures computed directly from the spike trains e.g., spike train correlations and dissimilarity metrics (DeAngelis et al. 1999; Escabí and Schreiner 2002; van Rossum 2001). Both the Elhilali study and our work demonstrate that the majority of STRFs change little across multiple recording sessions. Our conclusions are reached using a different technique for determining whether two STRFs derived from the same recording site are similar. Additionally, we incorporated similarity metrics that work at the spike-train level, demonstrating that there are correlations between STRF similarity and spike-train similarity measures.

Performance and Correlation-based Similarity

In addition to classifying stability using STRF-based similarity, we employed two more similarity measures: one based on neural performance and the other based on spike reliability correlations. Our lab recently explored these measures as they pertained to
individual recording performance. Specifically, one of our studies investigated the van Rossum spike distance metric and discussed its effectiveness as a model for multi- and single-unit spike discrimination of complex sounds (Narayan et al. 2006), while another studied the performances of various spike train classifiers, including the van Rossum SDM and the spike reliability correlation (Wang et al. 2007). Despite being used for spike classification, both measures were used because they achieve their results through different computations. We have previously shown that the performance-based computation and the spike reliability computations are similar analytically (Wang et al. 2007). Nevertheless, the two computations do vary in a number of ways: the spike reliability calculations are rate-normalized, while the performance-based metric is not; secondly, the spike reliability calculation is correlation-based, incorporating acausality, while the performance-based metric is causal; our implementation of the van Rossum metric calls for many iterations of the classification scheme to achieve an average result for the performance accuracy (Narayan et al. 2006), whereas the spike reliability calculation only runs through pairs of spike trains across recording blocks, averaging only across trials and stimulus responses. We were able to show, using both the neural performance measure and the spike reliability measure, that the sites display a wide range of variability in firing across experiments.

**STRF- Versus Spike Train-based Variability Measures**

We have seen that use of the different similarity measures yields different distributions of values. A greater fraction of STRF similarity values are allocated to the main peak of the histogram in Figure 1 than in Figure 2 and Figure 3. Figure 4 emphasizes this difference
more clearly; the STRF similarity index is more poorly correlated to either the $R_{\text{corr}}$ similarity measure or the performance-based similarity measure than these latter two are to each other (0.66 and 0.69 versus 0.95, respectively; $p < 0.001$ for all three).

Previous work investigating short-term plasticity has shown that understanding STRF variability is important to gauging how well a neuron will adapt to change induced by a tone detection task (Elhilali et al. 2007; Fritz et al. 2003). However, we have illustrated here that the sources of recording site variability may not be fully captured by the STRF. Our results – that the values of the variability measures based on the STRF are higher than those obtained using the $R_{\text{corr}}$ calculation or the performance-based calculations – may be due to one or two hypotheses. One hypothesis lies in the fact that, while the STRF is a reasonable estimate of those stimulus features that would best elicit a response in a neural unit, it represents a linear estimation of these features (Eggermont et al. 1983; Nelken et al. 1997; Theunissen et al. 2000). The variability measures defined using $R_{\text{corr}}$ and the distance metric may be able to account for the non-linearities that the STRF does not, and so these two additional measures complement the STRF-based variability measure, providing a more complete analysis of neural response variability. The second hypothesis for the higher values in the STRF-based variability measures relative to the $R_{\text{corr}}$- and performance-based measures may lie in the averaging involved in the calculations of these measures. The STRF is calculated using the peristimulus time histogram (PSTH), an average of the firing patterns of the recording site across the different trial presentation of the stimulus. On the other hand, both $R_{\text{corr}}$ and the performance-based similarity measure rely on spike train comparisons. As a result of
trial-to-trial variability, differences in spike alignment can lead to lower scores for the \( \text{\textit{R}}_{\text{corr}} \) and performance-based measures (Wang et al. 2007). Lower STRF similarity indices, therefore, may be the result of variable mean firing rates for the neural responses across recording blocks, while lower \( \text{\textit{R}}_{\text{corr}} \) and performance-based similarity measures may be due to inherent spike timing jitter and unreliable spikes in the neural responses at the individual recording block level (a hypothesis suggested by the correlations of the within-recording block measures to their respective across-recording block measures). We therefore suggest that, while measuring STRF variability is a necessary and logical step in understanding plasticity, the \( \text{\textit{R}}_{\text{corr}} \) and performance-based similarity measures serve to complement this description of variability found in neural recordings performed over multiple points in time.

\textit{Future Use of Similarity Measurements}

We have shown that the three measures used for this study quantify different ranges of variability based on the ways in which the calculations are performed. Further investigation into these measures could yield a benchmark that would define a range of variability for which future experiments involving plasticity would be viable; that is, given a unit's response to stimulus presentation over time, how likely is it that the unit will illustrate plastic behavior after being subject to a rapid-plasticity task?

Preliminary results from experiments in our laboratory suggest that, while it is possible to obtain task-free plasticity (unpublished observations) – that is, changes in STRF properties changing without the subject performing a behavioral task – the results of such
plasticity are highly variable. A task-based approach has been shown to be more robust in inducing more consistent plasticity results (Elhilali et al. 2007; Fritz et al. 2003). The results of our work here have helped us to assess the variability of awake neural recordings; a wide range of variability has been seen in the different sites from which the recordings were made. The hypotheses described above suggest a proper analysis of the potential sources of variability and how they may impact the results of a plasticity experiment. Although it may be tempting to determine that a more “stable” recording site – based on either of the stability measures or a combination of the three – would reveal a greater sensitivity to plastic changes that may not necessarily be the case. Variability primarily driven by within-recording block, trial-to-trial firing inaccuracy may weaken the ability to detect induced changes, since the inherent inaccuracy of the within-recording block neural responses would lead to decreased confidence in the observed responses. On the other hand, variation due to a slowly-changing mean neural response, represented by variability in the STRF, could be the target of a plasticity paradigm; the behavioral task’s demands could be tailored to affect this aspect of the neural response and would have a greater chance of producing observable, plastic changes in the properties of the recording site. We have also seen that some parameters of the STRF are more prone to change across recording blocks than others. Those properties that exhibit smaller changes would be ideal targets for tailoring specific plasticity experiments, as significant changes in these properties would be more easily detectable than in the properties that show high variability under control conditions.
ACKNOWLEDGEMENTS

We thank Rajiv Narayan and Elizabeth McClaine for help with the surgeries. This work was supported by National Institute on Deafness and Other Communication Disorders Grant 1R01 DC-007610-01A1.

REFERENCES


FIGURE 1. The majority of field L STRFs show high similarity across time as measured by the STRF similarity index. Panels (A) through (D) are STRF plots obtained from a particular neural recording site; timestamps at the upper-right corner of each STRF plot indicate the time elapsed after the start of the first recording block obtained at that same site. Warmer colors indicate excitatory regions and cooler colors inhibitory regions. (E) Plot of the STRF similarity index performed on all sites where multiple recordings were performed, plotted against the time elapsed after the start of the first recording block obtained at each site. The values of the similarity index for the site shown in panels (A) through (D) are outlined in red. To the right, a histogram shows the data in bins of size 0.05.

FIGURE 2. Illustration of spike correlation-based similarity (Rcorr) over time. Panels (A) through (D) are raster plots from the site whose STRFs are depicted in Figure 1(A-D); timestamps at the upper-right corner of each raster plot indicate the time elapsed after the start of the first recording block obtained at the same site. (E) Plot of across-recording block Rcorr where Rcorr was calculated between the responses from the first recording block and the responses from each subsequent recording block at a particular site plotted against the time elapsed after the start of the first recording block obtained at each site. To the right, a histogram shows the same data in bins of size 0.05. (F) Comparison of within-recording block Rcorr to across-recording block Rcorr, with a best-fit line plotted in black. The correlation coefficient was 0.98 (p < 0.001). In panels (E) and (F), the values of Rcorr for the data shown in panels (A) through (D) are outlined in red.

FIGURE 3. Discrimination-performance-based similarity shows relative stability over time at most sites. (A) Plot of across-recording block performance where the discrimination performance was calculated between the responses from the first recording block and the responses from each subsequent recording block at a particular site plotted against the time elapsed after the start of the first recording block obtained at the same site. To the right, a histogram shows the same data in bins of 5%. (B) Comparison of within-recording block Rcorr to across-recording block Rcorr, with a best-fit line plotted in black. The correlation coefficient was 0.95 (p < 0.001). The values of performance for the data shown in Figure 1(A-D) are outlined in red.

FIGURE 4. Rcorr and performance-based similarity are more correlated to each other than either is to the STRF similarity index. (A) Plot of across-recording block discrimination performance versus across-recording block Rcorr data. A best-fit line is plotted in black; the correlation coefficient between the two measures was 0.95 (p < 0.001). (B) Plot of across-recording block discrimination performance from Figure 3 versus the STRF similarity index. The correlation coefficient was 0.69 (p < 0.001). (C) Plot comparing across-recording block Rcorr data from Figure 2(E) to the STRF similarity index from Figure 1. The correlation coefficient was 0.66 (p < 0.001). The data from the example site in Figure 1 and Figure 2 are outlined in red.

FIGURE 5. Summary of STRF parameters over time. Each plot shows the changes in the STRF parameters described previously, where multiple recordings are made from each site, plotted against the elapsed time since the first recording at the same site. (A) Latency
(B) Center Frequency (CF) (C) Quality Factor (Q). (D) Temporal Integration Window (TiW). (E) Excitatory-Inhibitory Ratio (EIR) (F) Separability Index (SI). The data outlined in red correspond to the site featured in previous figures.

FIGURE 6. Similarity measures show low fractional change over time. Each set of connected symbols shows the fractional change in (A) Similarity Index values, (B) Across recording block $R_{corr}$ values, and (C) Performance-based similarity measure values, relative to the first value against the elapsed time since the first neural recording at a particular neural site. The red dashed lines indicate the standard error of the mean. As before, the example site is outlined in red.

SUPPLEMENTAL FIGURE 1. Example site showing high STRF variability. Panels (A) through (C) are STRF plots obtained from the neural recording site that illustrated the greatest variance in STRF similarity index; timestamps at the upper-right corner of each STRF plot indicate the time elapsed after the start of the first recording block obtained at that same site. Also indicated in the plots are the similarity indices calculated for the two corresponding STRFs.

SUPPLEMENTAL FIGURE 2. STRF signal-to-noise ratio is weakly correlated to STRF similarity index. The figure plots the STRF signal-to-noise ratio (SNR) in dB against the STRF similarity index, where the SNR values plotted correspond to the second and subsequent STRFs obtained at a particular recording site. SNR was calculated by dividing the power obtained in the causal region of the STRF (from 0 to 100 ms) by the power obtained in the acausal region of the STRF (from -150 to -50 ms). The correlation coefficient was 0.60 ($p < 0.001$).
Table 1. Percentage change in values of STRF parameters

<table>
<thead>
<tr>
<th>STRF Parameter</th>
<th>Average Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>3.3 ± 0.02</td>
</tr>
<tr>
<td>Center Frequency (CF)</td>
<td>10.9 ± 0.03</td>
</tr>
<tr>
<td>Quality Factor (Q)</td>
<td>17.1 ± 0.04</td>
</tr>
<tr>
<td>Temporal Integration Window (TiW)</td>
<td>7.2 ± 0.02</td>
</tr>
<tr>
<td>Excitatory-Inhibitory Ratio (EIR)</td>
<td>-0.3 ± 0.01</td>
</tr>
<tr>
<td>Separability Index (SI)</td>
<td>-1.0 ± 0.00</td>
</tr>
</tbody>
</table>

Values are mean ± SE. Average percentage change is calculated as the difference between the value of the STRF parameter obtained for the first neural recording at a site and all those obtained from subsequent recordings at that site, normalized by the value from the first recording, then averaged over all recordings and recording sites.

Table 2. Correlations among STRF parameters and variability measures

<table>
<thead>
<tr>
<th></th>
<th>CF</th>
<th>Q</th>
<th>TiW</th>
<th>EIR</th>
<th>SI</th>
<th>Sim Ind</th>
<th>R_corr</th>
<th>Perf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>0.36</td>
<td>0.30</td>
<td>0.69</td>
<td>-0.53</td>
<td>0.22</td>
<td>-0.46</td>
<td>-0.44</td>
<td>-0.48</td>
</tr>
<tr>
<td>CF</td>
<td>0.09</td>
<td>0.63</td>
<td>-0.24</td>
<td>0.03</td>
<td>0.59</td>
<td>-0.71</td>
<td>-0.60</td>
<td>-0.57</td>
</tr>
<tr>
<td>Q</td>
<td>0.53</td>
<td>-0.21</td>
<td>0.54</td>
<td>0.73</td>
<td>0.63</td>
<td>-0.32</td>
<td>-0.07</td>
<td>-0.13</td>
</tr>
<tr>
<td>TiW</td>
<td>-0.37</td>
<td>0.54</td>
<td>-0.59</td>
<td>0.22</td>
<td>0.20</td>
<td>-0.21</td>
<td>-0.04</td>
<td>-0.07</td>
</tr>
<tr>
<td>EIR</td>
<td>-0.06</td>
<td>0.12</td>
<td>0.22</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td></td>
<td>-0.21</td>
<td>-0.04</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations are as follows: CF = center frequency; Q = Quality Factor; TiW = Temporal Integration Window; EIR = Excitatory-Inhibitory Ratio; SI = Separability Index; Sim Ind = Similarity Index; Perf = Performance-based similarity measure. Values in italics indicate significant correlations ($p < 0.001$).
A. 0 min
B. 29 min
C. 1 hour 42 min
D. 3 hours 7 min
E. STRF Similarity Index vs. Elapsed Time (min)
Across-Recording Block Performance (% Correct)

Elapsed Time (min)

% of Recordings

Across-Recording Block Performance (% Correct)

Within-Recording Block % Correct