Neuron-Specific Stimulus Masking Reveals Interference in Spike Timing at the Cortical Level

ERIC LARSON^{1,2}, ROSS K. MADDOX^{1,2}, BEN P. PERRONE^{1,2}, KAMAL SEN^{1,2}, AND CYRUS P. BILLIMORIA^{1,2}

¹*Department of Biomedical Engineering, Hearing Research Center, Boston University, 44 Cummington St, Boston, MA 02215, USA*

²*Center for Biodynamics, Boston University, 44 Cummington St, Boston, MA 02215, USA*

Received: 27 April 2011; Accepted: 8 September 2011; Online publication: 1 October 2011

ABSTRACT

The auditory system is capable of robust recognition of sounds in the presence of competing maskers (e.g., other voices or background music). This capability arises despite the fact that masking stimuli can disrupt neural responses at the cortical level. Since the origins of such interference effects remain unknown, in this study, we work to identify and quantify neural interference effects that originate due to masking occurring within and outside receptive fields of neurons. We record from single and multi-unit auditory sites from field L, the auditory cortex homologue in zebra finches. We use a novel method called spike timing-based stimulus filtering that uses the measured response of each neuron to create an individualized stimulus set. In contrast to previous adaptive experimental approaches, which have typically focused on the average firing rate, this method uses the complete pattern of neural responses, including spike timing information, in the calculation of the receptive field. When we generate and present novel stimuli for each neuron that mask the regions within the receptive field, we find that the time-varying information in the neural responses is disrupted, degrading neural discrimination performance and decreasing spike timing reliability and sparseness. We also find that, while removing stimulus energy from frequency regions outside the receptive field does not significantly affect neural responses for many sites, adding a masker in these frequency regions can nonetheless have a signifi-

cant impact on neural responses and discriminability without a significant change in the average firing rate. These findings suggest that maskers can interfere with neural responses by disrupting stimulus timing information with power either within or outside the receptive fields of neurons.

Keywords: masking, birdsong, auditory cortex, spectrotemporal receptive field, neural coding

INTRODUCTION

The near silence of an auditory neuroscience lab is an atypical environment for listening to sounds. Usually, humans must listen to a target sound source (such as a person talking) in the presence of background maskers (e.g., other talkers, traffic noise, a loud TV, etc.). Many animals, such as birds (Hulse et al. 1997), frogs (Endepols et al. 2003), and mammals (Cherry 1953), are capable of listening to a single sound source in the presence of masking sounds. Although this ability is important for both humans and animals, how the brain performs this task remains unclear, and understanding the underlying neural mechanisms might be critical for developing new strategies for hearing devices (Haykin and Chen 2005).

Previous studies suggest an important role for auditory cortex in processing sounds in the presence of maskers (Nelken 2004). Here we examined the neural processing of masked stimuli at the cortical level in songbirds. The songbird system is characterized by a combination of well-understood vocal communication behaviors and identified neural circuits that mediate the perception, learning, and production of vocal communication sounds. Moreover, songbirds communicate in crowded and noisy colonies, making

Eric Larson and Ross K. Maddox are first authors who contributed equally to this work.

Correspondence to: Cyrus P. Billimoria · Department of Biomedical Engineering, Hearing Research Center · Boston University · 44 Cummington St, Boston, MA 02215, USA. Telephone: +1-617-3539148; fax: +1-617-3536766; email: cpb@bu.edu

them an attractive model system for studying the neural processing of masked sounds. Understanding the neural processing of masked stimuli in songbirds, then, could help us understand how the brain manages to recognize target stimuli embedded in noise. Previously, we identified different forms of neural interference effects that lead to a dramatic reduction in discrimination of target sounds in the presence of background maskers (Narayan et al. 2007), where masking caused additions of spurious spiking during gaps in songs and the removal of informative spiking during song syllables. However, the origin of such neural interference effects remains unknown.

In this study, we aim to characterize the origin of this neural interference using a new adaptive stimulation method called spike timing-based stimulus filtering (STSF). Based on the neural response, including spike timing information, the STSF method calculates a receptive field estimate and creates new stimuli for each recording site, which allow us to place maskers in different regions relative to the receptive field. We utilize this method to examine neural responses to masked target birdsongs in field L, the avian auditory cortex homologue (Wang et al. 2010), as field L has been shown to respond more strongly to complex stimuli such as conspecific vocalizations than tones (Leppelsack and Vogt 1976; Langner et al. 1981), noises, or synthetic stimuli (Grace et al. 2003). Field L has also been shown to contain sufficient information to classify different birdsongs on the basis of responses from single neurons (Wang et al. 2007) and provides input to downstream areas that show song-selective responses (Gentner and Margoliash 2003). Using this method, our results reveal interference with the neural response and disruption of coding of target identity when maskers are within the receptive field, as well as when maskers are placed outside the receptive field, suggesting different ways the coding of stimulus identity can be disrupted by the presence of maskers.

MATERIALS AND METHODS

Electrophysiological recording

All procedures were in accordance with the National Institutes of Health guidelines approved by the Boston University Institutional Animal Care and Use Committee. Single-electrode extracellular neural responses from field L in adult male zebra finches (*Taeniopygia guttata*) were recorded using previously developed electrophysiological techniques for acute (Narayan et al. 2006; Billimoria et al. 2008) and awake-restrained (Grana et al. 2009) recordings. We used 2–4 M Ω tungsten microelectrodes and presented conspecific vocalizations at 72 dB sound

pressure level (SPL) to probe for auditory sites. Subsequently filtered and processed stimuli varied in level from 53.3 to 74.5 dB SPL (see “[Spike timing-based stimulus filtering](#)” below). Sites with audible time-locked neural activity and sufficient van Rossum discrimination performance (>70%) on two targets (see “[Spike train analysis](#)” below) were isolated using threshold-based spike detection and classified as single or multi-unit based on the percentage of inter-spike interval (ISI) violations (criterion—1 ms ISI violations less than 5% were considered single units). We used stereotactic coordinates to evenly sample field L, but did not obtain sufficient data to draw conclusions about the effects of subregion on neural responses.

Spike timing-based stimulus filtering

Once sites were identified, 20 different zebra finch songs (total duration of 40.5 s) were each played 10 times in pseudorandom block order (Fig. 1A). Using the responses to these stimuli, the receptive field (spectrotemporal receptive field, STRF; Fig. 1B) for each site was estimated using normalized reverse correlation (NRC) with STRFPak 5.3 (Theunissen et al. 2001), which uses time-varying firing rates to estimate the neural receptive field. Zebra finch songs were used to calculate the STRF instead of continuous, wideband stimulation because field L neurons respond more strongly to birdsong than noise (Grace et al. 2003), and songs would be used as target stimuli in subsequent analyses. For the STRF estimate, 64 frequency bands spanning the range from 250 to 8,000 Hz were used with 2 ms temporal binning of the spectrogram (log power in each time-frequency bin). These STRF estimates were cross-validated using STRFPak to calculate the mutual information and the noise-corrected cross-correlation (CC) between the STRF-predicted response and actual responses using a leave-one-out jackknifing procedure (Hsu et al. 2004).

We then implemented the STSF method using custom software. To reduce experimental time, 5 of the original 20 songs used to calculate the RF were selected to be targets in constructing an individualized set of stimuli for each recording site by spectrally filtering the targets according to each site’s RF. For each frequency band, the contribution to the receptive field was computed as the maximum positive (excitatory) value plus the magnitude of the minimum negative (inhibitory) value across time in the RF (Fig. 1C). Considering the extrema across time to calculate the contribution helped account for the fact that temporal interactions across frequencies could affect the neural responses. Using this across-time contribution measure allowed for a simple stimulus filtering scheme. Frequencies whose contributions

FIG. 1. After STRF calculation, the contribution of each frequency band to the receptive field is estimated. **A** Five example song spectrograms (power in each frequency versus time) are shown which elicited example field L responses (rasters). **B** Responses to 20 songs were used to generate the STRF, which estimated the stimulus features that increased (red)

and decreased (blue) the firing probability of the example site. **C** The contribution of each frequency band (black line) was calculated by adding the minimum (inhibitory, blue) and maximum (excitatory, red) magnitude in each frequency band. Different receptive field inclusion thresholds (25%, 50%, and 75%) were calculated relative to the maximum contribution.

were less than a threshold percentage (at 25%, 50%, or 75%) of the maximal contribution across all frequency bands (Fig. 1C) were filtered out to produce stimuli where the target was within the NRC-estimated receptive field (T_w) from the target stimuli (T). For example, at the 25% threshold, frequencies whose contributions were less than 25% of the maximal contribution (across frequency bands) were removed via filtering (for filtering details, see below).

For each site, we chose the highest threshold percentage that maintained neural responses between the unfiltered target stimuli (T) and the within-STRF stimuli (T_w). To quantify the neural response changes, we utilized a spike train discrimination method (Machens et al. 2003). This procedure used a nearest-neighbor template-matching procedure to classify which spike trains were evoked in response to different stimuli. Under this procedure, to classify one spike train evoked in response to a particular song, the spike train was compared to randomly chosen template spike trains evoked in response to each song and classified according to the smallest dissimilarity to the template spike trains. Repeating this over each trial from each song yielded a percent correct discrimination score. Spike train dissimilarities were calculated using the van Rossum (2001) metric as the integral of the squared difference between spike trains temporally smoothed with a decaying exponential (time constant τ ; for each site, τ was chosen to maximize discrimination performance for the unmodified T stimuli, spanning a range for our data

of 0.5 to 70 ms). We first used this discrimination method to determine baseline performance by discriminating neural responses to T stimuli using the responses to T stimuli as templates. To quantify neural response changes due to within-STRF filtering, at each filtering threshold (25%, 50%, and 75% of the maximal contribution) we then discriminated responses to T_w stimuli using responses to T stimuli as templates. Finally, for each site, we chose the highest filtering threshold that yielded less than a 5% change in discrimination performance (compared to baseline performance) due to filtering. Frequency bands whose contributions were above the threshold were termed “within-STRF” and bands whose contribution was below were termed “outside-STRF.” By selecting the threshold in this manner, we obtained conservative estimates of which frequencies were labeled outside-STRF.

Stimulus filtering was performed using 64 parallel 5,000-point finite-impulse-response filters with delay correction. In the frequency domain, each filter was a triangle that obtained unity magnitude at the center frequency of a band, dropping linearly to zero at the center frequencies of the adjacent bands and taking on the value of zero elsewhere. Filtering via this time domain method prevented both spectral splatter and temporal ringing (introducing at most 113 ms of ringing, but far less in practice). Although this filtering did result in minor spectral overlap between adjacent bands, this overlap was mitigated by the smoothness of STRF estimates that arises from the singular value decomposition used in normalized

reverse correlation calculations (David et al. 2007). Removal of the outside-STRF frequencies resulted in filtered target within-STRF (Tw) stimuli with sound levels between 53.3 and 71.9 dB SPL.

In addition to the target stimuli, a random-phase noise masker was generated whose spectrum matched the average of the conspecific stimuli used to generate the STRF, but that contained no temporal structure that could be used for discriminating songs. Since the overall spectrum of the noise matched that of multiple birdsongs, this noise masker can be thought of as a surrogate for the sound from a very crowded bird colony environment. After creating the target within-STRF stimuli (Tw) and establishing the inclusion threshold, this noise masker was added to the Tw stimuli in all within-STRF frequency bands (TwMw) or outside-STRF bands (TwMo) (Fig. 2). In each outside-STRF band, then, the randomized masker level matched the average level in that band across the original stimuli. For each site, all stimuli (5 targets with T, Tw, TwMw, and TwMo variations) were presented in pseudorandom order 10 times, with a different token of noise masker (i.e., unfrozen noise) used for each of the 10 trials. Unfrozen noise was used here because noise maskers typically vary between repeated presentations of target stimuli in listening environments. All stimuli were truncated to the

duration of the shortest target song (820 ms) and all but the shortest song were trimmed of introductory notes (which carry little information about song identity). The resulting stimuli had sound levels between 56.9 and 74.5 dB SPL. The duration of the presentation of all stimuli varied by site, typically lasting less than an hour (longest was 82 min).

Spike train analysis

To quantify changes in spike trains due to stimulus modification, we used the previously described van Rossum discrimination method to discriminate neural responses to the modified stimuli using the unmodified-target (T) spike trains as templates. Although it is unknown how the zebra finch auditory system allows for behavioral discrimination of songs, the van Rossum discrimination measure helps quantify how much information is available in the spike timing information for subsequent discrimination by downstream neurons. We also measured the reliability and sparseness of neural responses using previously developed techniques. We used R_{corr} (Schreiber et al. 2003) to measure reliability of the neural responses. This is obtained by first calculating the mean correlation between all pairs of trials of Gaussian-smoothed spike trains evoked in response to the same song (yielding values between 0

FIG. 2. The STSF method generates and tests novel stimuli during the experiment based upon the contribution of frequencies to the receptive field. **A** The STRF (left) for an example site and two of the five target stimuli (spectrograms) are shown, along with neural responses (rasters). **B** When target stimuli were filtered at the 50% threshold to remove frequencies outside the RF (spectrograms; Tw),

neural responses were similar to the unfiltered responses. **C** Adding a masker in the frequency bands within the STRF to the filtered stimuli (spectrograms; target within STRF, masker within STRF: TwMw) elicited different neural responses (rasters). **D** Adding a noise masker to the frequency bands outside the RF (spectrograms; target within STRF, masker outside STRF: TwMo) also degraded neural response timing.

and 1) and then averaging these values across all songs. We measured sparseness (Vinje and Gallant 2000) by using a PSTH-binning technique to determine how concentrated in time neural responses were (with values between 0 and 1). We used these measures to help quantify characteristics of the time neural responses to stimuli. Time constants for calculating sparseness and reliability were matched to the optimal neural time scale from the discrimination measure. We also measured the overall spike rate for each site despite the fact that songs are not readily discriminable using spike rate (Narayan et al. 2006; Larson et al. 2009).

To establish significant differences across stimulus conditions (T, Tw, TwMw, and TwMo), we used one-way repeated measures ANOVAs for parametric data that passed Kolmogorov–Smirnov normality and Levene’s equal variance tests; data that did not pass these tests were analyzed using Friedman’s nonparametric repeated measures test. Post hoc multiple pairwise comparisons were performed using Tukey’s honestly significant difference test. All tests were performed at a $p < 0.05$ significance level and are shown in figures by [*, **, ***] at the [0.05, 0.01, 0.001] levels, respectively.

RESULTS

Neurophysiology and spike timing-based stimulus filtering

We recorded extracellular responses from 34 sites in field L, the mammalian auditory cortex homologue in zebra finches. Thirty-three sites were recorded from 6 anesthetized birds, and one site was recorded in an awake-restrained bird. Of these, nine sites were classified as single unit recordings (1 ms ISI violations less than 5%). Since we did not observe significant differences between the single unit and multi-unit recordings in terms of their linearity (as measured by the CC and predicted information [Hsu et al. 2004], $p > 0.53$ each, unpaired t test), they were combined for subsequent analysis. For each site, 20 conspecific songs were presented (5 shown in Fig. 1A), and neural responses were used to calculate the spectro-temporal receptive field (STRF) using normalized reverse correlation (Fig. 1B).

To test the effects of masking on the neural responses, we used these receptive field estimates to generate new stimuli for each neural recording site. To do this, we first determined the contribution of each frequency band to the receptive field estimate (Fig. 1C) and used that to filter out frequencies deemed outside the STRF from five target songs. For each site, a frequency band contribution threshold was chosen (at 25%, 50%, or 75% of the maximal contribution across frequency bands) that preserved

the time-varying neural responses, as measured by the van Rossum discrimination method (see “Materials and methods”). We then created site-specific within-STRF-only (Tw) stimuli, and the target (T) stimuli were filtered to get rid of frequencies with contributions below the threshold chosen for each site. There were nine sites whose neural responses changed (according to our threshold-determining criterion) while filtering out frequencies with contributions below the lowest threshold (25%), and these were not included in the subsequent site-specific filtering analysis as their estimate of the receptive field was considered inadequate for site-specific stimulus filtering. The remaining 25 sites had STRF CC values that ranged from 0.48 to 0.88 ($\mu = 0.64$), and these CC values did not differ significantly from those of the disqualified sites ($\mu = 0.60$, $p = 0.11$, unpaired t test). However, the relative mean predicted information values did significantly differ ($p = 0.02$, unpaired t test), with the mean predicted information for included sites (19.33 bits/s) greater than that for the disqualified sites (7.62 bits/s).

Of the 25 sites that met the inclusion criterion, 9, 11, and 5 sites used 25%, 50%, and 75% contribution thresholds, respectively, to delineate the within- and outside-STRF regions. Using these thresholds, we created four classes of stimuli using the target songs and random noise maskers spectrally matched to zebra finch song designed to determine the effects of within-STRF and outside-STRF frequency masking effects for each neural site. These stimuli were unfiltered target (T), within-STRF-only target (Tw), within-STRF target plus within-STRF masker (TwMw), and within-STRF target plus outside-STRF masker (TwMo) (Fig. 2).

Contribution-based frequency filtering can preserve neural responses

In filtering out the frequency bands outside the RF (Tw) from the full stimuli (T), neural response properties were preserved across sites despite significant changes in overall stimulus intensity (Fig. 3A). To quantify how the neuron’s response timing and reliability changed, we performed van Rossum-based discrimination of the spike trains evoked in response to the filtered songs while using spike trains from the unfiltered (T) stimuli as templates (Fig. 3B). We also measured the reliability of the neural responses themselves (Fig. 3C)—calculated as the average correlation between pairs of temporally smoothed spike trains evoked in response to the same song. There was no significant change in the discrimination, reliability, sparseness (Fig. 3D), or firing rate (Fig. 3E) of the neural responses due to filtering out the frequencies outside the STRF. This suggests that the timing of neural responses to the stimuli was not

FIG. 3. Frequencies outside the receptive field can be removed using site-specific filtering while preserving neural responses, and masking frequencies within the STRF disrupts neural responses. **A–E** When target stimuli (T) were filtered to contain only within-STRF frequency bands (Tw), there were significant changes in the stimulus intensity (**A**) but no significant change in the neural discriminability (**B**), spike timing

reliability (**C**), sparseness (**D**), or firing rate (**E**), suggesting that the timing of neural responses was predominantly preserved by filtering. Once a masker was added to these within-STRF frequency bands (TwMw), the discriminability, reliability, and sparseness were all affected. Individual sites are gray, means (± 1 SEM) in black; two outlier (very high rate) sites are omitted from the firing rate plot but included in the mean.

affected by filtering out the outside-STRF frequencies using the chosen contribution threshold.

Masking frequency bands within the receptive field degrades responses

When a noise masker was added to the frequency bands within the STRF with the target (TwMw stimuli), the neural responses changed across all sites. We found no significant change in overall firing rate across sites. We then measured the neural discriminability (see “Materials and methods”), which quantified how well the neural responses to these modified stimuli could be correctly classified based on their similarity to template spike trains evoked in response to the unmodified song stimuli. We observed significant changes in the mean discriminability, as well as the reliability and sparseness of the neural responses (Fig. 3). This suggests that, despite similar overall firing rates, time-varying neural response properties changed—and these changes adversely affected the ability to discriminate songs on the basis of the neural response.

Effects of masking frequency bands outside the receptive field

When the noise masker was placed in the frequency bands outside the STRF with the target in the frequency bands within the STRF (TwMo stimuli), the time-varying response properties of sites also changed (Fig. 4A) with significant decreases in discrimination and spike timing reliability (Fig. 4B, C). That is, despite the fact that removing outside-STRF frequencies from the stimuli did not significantly change discriminability or reliability (Fig. 4B, C, $p > 0.47$ for both), adding a noise masker to outside-STRF frequency regions decreased discriminability ($p < 0.001$), and this decrease was accompanied by a corresponding decrease in spike timing reliability ($p < 0.001$). However, the overall firing

rate (Fig. 4E) and sparseness of the responses (Fig. 4D) did not significantly change ($p > 0.18$ for both). In addition, we observed that there was a correlation between the number of outside-STRF bands and site linearity (as measured by the predicted information provided by the STRF; $R = 0.68$, $p < 0.001$). This suggests that the units better described by the linear model could have more frequency content removed without affecting the neural responses.

DISCUSSION

Within and outside receptive field effects in degradation of discrimination

Despite the importance of the recognizing sounds in the presence of maskers, the neural processing underpinning this ability is not fully understood, in part because of the lack of studies examining neural response properties with masking stimuli. Although disruptions in neural timing have been observed (Narayan et al. 2007), the underlying sources of these effects are not known. We sought to determine if interference effects originated from masking within the receptive field, or if some effects were due to masking stimuli outside the receptive field.

We found that adding a masker to our stimuli in the frequency bands within the receptive fields of individual sites disrupted neural coding of song identity despite no significant changes in mean firing rate. Discrimination degraded because the spike timing in response to the target sound was disrupted by the presence of the masker. This decrease in performance was accompanied by a significant decrease in spike timing reliability and sparseness of the neural responses. This suggests an improvement in neural coding with sparseness, consistent with experimental studies in the visual system (Vinje and Gallant 2000). The observed decreases in spike timing reliability suggest that intro-

FIG. 4. Adding a masker in frequency bands outside the receptive field can disrupt neural responses. **A** Three example field L sites' spike trains change in response to site-specific filtered stimuli (spectrograms above each raster). The response from each site is preserved in filtering to remove outside-STRF frequencies ($T \rightarrow Tw$), but degrades in response to a masker placed in those outside-STRF frequency regions ($TwMo$). **B, C** The neural discrimination (**B**) and the spike timing reliability (**C**) of responses

significantly changed relative to the filtered stimuli (Tw) due to the addition of a noise masker to the frequency bands outside the STRF ($TwMo$). **D, E** The sparseness (**D**) and firing rate (**E**) for each site are shown; individual traces for two outlier (high rate) sites are not shown on the firing rate plot. **B–E** Individual sites (gray), plus sites 1–3 (red, blue, green) and the site from Figures 1 and 2 (orange), with means (± 1 SEM) in black.

ducing the noise masker decreased the trial-to-trial reliability of neural coding of the target stimuli in the presence of maskers that vary across target presentations. It is possible that some of the neural response was being driven by the noise masker in a reliable manner, and reliability decreased because the specific masker noise token varied from trial to trial; however, this is somewhat unlikely as previous experiments have observed little phase locking to band-passed white noise stimuli in field L (Grace et al. 2003).

We also found that, for roughly three quarters of the sites tested, filtering out frequencies that did not contribute to the STRF estimate did not affect the time-varying response properties of neural responses. This suggests that these neurons were more linear in their stimulus–response characteristics than the remaining quarter. However, even for these predominantly linear neurons, we observed a range of effects on responses due to the addition of a spectrally matched noise masker in the outside-STRF frequency bands. For some sites, neural responses—as measured by discrimination, reliability, firing rate, and sparseness—were preserved, while other sites showed large changes in their responses, often

manifesting as degradation in the neural coding of song identity as measured by the discrimination performance. Although it is unclear why this degradation occurs, it could be caused by the presence of subthreshold inputs from neurons tuned to these outside-STRF frequency regions, or superthreshold inputs that contribute minimally to the STRF estimate. Also, while here we have termed certain frequencies “outside” the STRF based on responses to restricted bandwidth target stimuli, this is an artificial definition. It is clear that there are interactions in the frequency regions that initially appear to contribute little to the neural response, but it could be informative to test other configurations to probe these interactions. For example, we did not test neural responses to target stimuli placed in the frequency bands that we termed “outside” the receptive field; testing additional stimuli such as these could help clarify how different regions of the receptive field affect neural responses.

The STSF method

In this study, we used a novel adaptive stimulation paradigm called STSF. Previously, adaptive stimulation

paradigms have used tone stimuli to maximize neural firing rate (deCharms et al. 1998); determined receptive fields in response to three-dimensional visual stimuli in macaque inferotemporal cortex (Yamane et al. 2008); generated nonlinear interaction maps for firing rates in mammalian auditory cortex (Barbour and Wang 2003; Sadagopan and Wang 2009) and bat inferior colliculus (Brimjoin and O'Neill 2010) using multiple tone stimuli; characterized how field L firing rates change with intensity using linear–nonlinear models (Nagel and Doupe 2006); and shown that cortical neurons change firing rate in response to removal of background noise or other stimulus modifications (Bar-Yosef et al. 2002). Although these previous studies have successfully extended characterizations of neural response properties in terms of firing rates, little work has been done to examine how stimulus changes alter the timing of cortical responses; so far, it has been shown that the timing of cortical responses to foreground stimuli can be modulated by the acoustic background (Bar-Yosef and Nelken 2007). By examining in detail the changes in neural response timing—in addition to changes in overall firing rate—due to site-specific stimulus modifications, we can obtain a clearer picture of neural response properties, and how these properties may contribute to effective downstream processing, such as stimulus discrimination and recognition.

The adaptive stimulation method outlined here builds upon those previous adaptive stimulation paradigms, extending previous work in three important ways. First, we used complex natural communication sounds to which field L preferentially responds (Grace et al. 2003; Theunissen and Shaevitz 2006) instead of synthetic stimuli. Second, we utilized both frequency complementary and overlapping noise maskers in the novel adaptive stimulus set to probe for within and outside receptive field effects, respectively. Third, and most importantly, we analyzed changes in the precise spike timing. This revealed differences in neural response characteristics despite the lack of significant changes in overall firing rate.

Because receptive field estimation can be challenging, our method uses the STRF only as a starting point for generating novel stimuli and collecting neural responses. For example, it is known that STRFs are generally not good predictors of neural responses for stimuli with different spectrotemporal properties from those used to calculate the STRF (Theunissen et al. 2000; Christianson et al. 2008; Gourevitch et al. 2009), so here we used the STRF only to estimate which frequencies mattered to each neuron in the relevant discrimination task stimuli (birdsong). Additionally, normalized reverse correlation imposes spectral smoothness on the STRF estimates (David et al. 2007), and the resulting overestimation of

the range of important frequencies allowed us to be conservative in our frequency removal procedure. As a final precaution, we also removed from consideration roughly one quarter of the units, those that the STRF did not predict well. In the future, it could prove useful to take a similar approach with other receptive field estimates, such as boosting (David et al. 2007), generalized linear models (Paninski et al. 2007), or multilinear models (Ahrens et al. 2008). The STSF method introduced here can be used as a tool to explore differences between receptive field estimates and to test other cortical neuron models.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders Grant 1R01 DC-007610-01A1 and a National Science Foundation Graduate Research Fellowship awarded to Eric Larson.

REFERENCES

- AHRENS MB, LINDEN JF, SAHANI M (2008) Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci* 28:1929–1942
- BARBOUR DL, WANG X (2003) Contrast tuning in auditory cortex. *Science* 299:1073–1075
- BAR-YOSEF O, NELKEN I (2007) The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Front Comput Neurosci* 1:3
- BAR-YOSEF O, ROTMAN Y, NELKEN I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J Neurosci* 22:8619–8632
- BILLIMORIA CP, KRAUS BJ, NARAYAN R, MADDOX RK, SEN K (2008) Invariance and sensitivity to intensity in neural discrimination of natural sounds. *J Neurosci* 28:6304–6308
- BRIMJOIN WO, O'NEILL WE (2010) Patterned tone sequences reveal non-linear interactions in auditory spectrotemporal receptive fields in the inferior colliculus. *Hear Res* 267:96–110
- CHERRY EC (1953) Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25:975–979
- CHRISTIANSON GB, SAHANI M, LINDEN JF (2008) The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J Neurosci* 28:446–455
- DAVID SV, MESGARANI N, SHAMMA SA (2007) Estimating sparse spectrotemporal receptive fields with natural stimuli. *Network* 18:191–212
- DECHARMS RC, BLAKE DT, MERZENICH MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443
- ENDEPOLIS H, FENG AS, GERHARDT HC, SCHUL J, WALKOWIAK W (2003) Roles of the auditory midbrain and thalamus in selective phonotaxis in female gray treefrogs (*Hyla versicolor*). *Behavioural Brain Research* 145:63–77
- GENTNER TQ, MARGOLIASH D (2003) Neuronal populations and single cells representing learned auditory objects. *Nature* 424:669–674
- GOUREVITCH B, NORENA A, SHAW G, EGGERMONT JJ (2009) Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. *Cereb Cortex* 19:1448–1461

- GRACE JA, AMIN N, SINGH NC, THEUNISSEN FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. *J Neurophysiol* 89:472–487
- GRANA GD, BILLIMORIA CP, SEN K (2009) Analyzing variability in neural responses to complex natural sounds in the awake songbird. *J Neurophysiol* 101:3147–3157
- HAYKIN S, CHEN Z (2005) The cocktail party problem. *Neural Computation* 17:1875–1902
- HSU A, BORST A, THEUNISSEN FE (2004) Quantifying variability in neural responses and its application for the validation of model predictions. *Network* 15:91–109
- HULSE SH, MACDOUGALL-SHACKLETON SA, WISNIEWSKI AB (1997) Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *J Comp Psychol* 111:3–13
- LANGNER G, BONKE D, SCHEICH H (1981) Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Exp Brain Res* 43:11–24
- LARSON E, BILLIMORIA CP, SEN K (2009) A biologically plausible computational model for auditory object recognition. *J Neurophysiol* 101:323–331
- LEPPELSACK HJ, VOGT M (1976) Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *J Comp Physiol* 107:263–274
- MACHENS CK, SCHUTZE H, FRANZ A, KOLESNIKOVA O, STEMMLER MB, RONACHER B, HERZ AV (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci* 6:341–342
- NAGEL KI, DOUPE AJ (2006) Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* 51:845–859
- NARAYAN R, GRANA G, SEN K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96:252–258
- NARAYAN R, BEST V, OZMERAL E, MCCLAIN E, DENT M, SHINN-CUNNINGHAM B, SEN K (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci* 10:1601–1607
- NELKEN I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol* 14:474–480
- PANINSKI L, PILLOW J, LEWIS J (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. In: Cisek P, Drew T, Kalaska J (eds) *Computational neuroscience: theoretical insights into brain function*. Elsevier, Amsterdam, pp 493–507
- SADAGOPAN S, WANG X (2009) Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci* 29:11192–11202
- SCHREIBER S, FELLOUS JM, WHITMER D, TIESINGA P, SEJNOWSKI TJ (2003) A new correlation-based measure of spike timing reliability. *Neurocomputing* 52:925–931
- THEUNISSEN FE, DAVID SV, SINGH NC, HSU A, VINJE WE, GALLANT JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316
- THEUNISSEN FE, SEN K, DOUPE AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331
- THEUNISSEN FE, SHAEVITZ SS (2006) Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol* 16:400–407
- VAN ROSSUM MC (2001) A novel spike distance. *Neural Comput* 13:751–763
- VINJE WE, GALLANT JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–1276
- WANG L, NARAYAN R, GRANA G, SHAMIR M, SEN K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27:582–589
- WANG Y, BRZOSOWSKA-PRECHTL A, KARTEN HJ (2010) Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences* 107:12676–12681
- YAMANE Y, CARLSON ET, BOWMAN KC, WANG Z, CONNOR CE (2008) A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11:1352–1360