Journal of Computational Neuroscience 17, 47–56, 2004

Searching for Optimal Sensory Signals: Iterative Stimulus Reconstruction in Closed-Loop Experiments

FREDRIK EDIN,* CHRISTIAN K. MACHENS,† HARTMUT SCHÜTZE AND ANDREAS V.M. HERZ**
Institute for Theoretical Biology, Humboldt Universität zu Berlin, Invalidenstr. 43, 10115 Berlin, Germany
a.herz@biologie.hu-berlin.de

Received June 19, 2002; Revised November 30, 2003; Accepted February 19, 2004

Action Editor: Israel Nelken

Abstract. Shaped by evolutionary processes, sensory systems often represent behaviorally relevant stimuli with higher fidelity than other stimuli. The stimulus dependence of neural reliability could therefore provide an important clue in a search for relevant sensory signals. We explore this relation and introduce a novel iterative algorithm that allows one to find stimuli that are reliably represented by the sensory system under study. To assess the quality of a neural representation, we use stimulus reconstruction methods. The algorithm starts with the presentation of an initial stimulus (e.g. white noise). The evoked spike train is recorded and used to reconstruct the stimulus online. Within a closed-loop setup, this reconstruction is then played back to the sensory system. Iterating this procedure, the newly generated stimuli can be better and better reconstructed. We demonstrate the feasibility of this method by applying it to auditory receptor neurons in locusts. Our data show that the optimal stimuli often exhibit pronounced sub-threshold periods that are interrupted by short, yet intense pulses. Similar results are obtained for simple model neurons and suggest that these stimuli are encoded with high reliability by a large class of neurons.

Keywords: neural coding, stimulus reconstruction, insect, auditory receptor

1. Introduction

During the last decade, our understanding of the neural code has been significantly advanced by stimulus reconstruction methods (Bialek et al., 1991; Gabbiani et al., 1996; Theunissen et al., 1996; Stanley et al., 1999; Machens et al., 2001). Within this framework, one seeks to reconstruct a stimulus or some of its features from the neural activity that it evokes. The success of the method depends not only on a proper reconstruction algorithm but also on various stimulus characteristics such as bandwidth and variance (Wessel et al., 1996). For instance, the reconstruction of natural stimuli or artificial stimuli tailored to the specific properties of a sensory system often works better than the reconstruction of more general stimuli such as white noise (Rieke et al., 1995; Machens et al., 2001).

Here we seek to investigate the distinctive features of stimuli that can be well reconstructed. For that purpose, we use an algorithm that iteratively approaches such stimuli during the course of an experiment. In each iteration we reconstruct a given stimulus online and play the reconstruction back to the sensory system under study. This procedure eliminates step by step any stimulus features that cannot be well reconstructed. Hence, we obtain a series of stimuli and spike trains that are more and more robust against perturbations by internal neural noise sources. From an evolutionary point of

*Present address: Department of Numerical Analysis and Computer Science, Royal Institute of Technology, SE-100 44, Stockholm, Sweden.
†Present address: Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY 11724, USA.
**To whom correspondence should be addressed.

view, such stimuli are likely to share statistical properties of behaviorally relevant sensory signals as those are supposedly encoded more faithfully.

In particular, we are interested in three questions: does the reconstruction series converge towards non-trivial stimuli that can be reconstructed almost perfectly? What distinguishes the stimulus/spike-train combinations that are most robust against noise? What do these combinations tell us about the functional properties of the particular neuron under investigation?

The algorithm relies on the prior availability of a successful reconstruction method. Here we test the procedure on auditory receptor neurons of locusts, a system that has previously been demonstrated to be well accessible to linear stimulus reconstruction methods (Machens et al., 2001), and support our results with numerical studies on integrate-and-fire model neurons.

2. Materials and Methods

2.1. Iterative Stimulus Reconstruction

Linear stimulus reconstruction methods (Bialek et al., 1991; Rieke et al., 1997) provide a way to estimate a stimulus from the evoked spike train. The reconstructed stimulus can only contain stimulus features encoded in the spike train. Aspects of the stimulus that cannot be reconstructed include features that are not encoded by the specific neuron, features that have been lost due to various noise sources, and features that cannot be reconstructed with linear methods. A search for stimuli that can be reliably reconstructed could thus start with a rather general stimulus and subsequently eliminate all those parts that cannot be reconstructed. This might be achievable by the following iterative procedure (cf. Fig. 1):

(1) Start with an initial, general stimulus $s_1(t)$, e.g. Gaussian white noise.

At the n-th iteration:

(2) Present stimulus $s_n(t)$ and record spike train $y_n(t)$.
(3) Compute an estimate $s_n^{\text{est}}(t)$ of the stimulus $s_n(t)$,

$$s_n^{\text{est}}(t) = h_n + \int_0^T d\tau k_n(\tau) y_n(t-\tau) \qquad (1)$$

where the parameter h_n and the kernel $k_n(\tau)$ are chosen so as to minimize the mean-square difference $\langle (s_n^{\text{est}}(t) - s_n(t))^2 \rangle$ between stimulus and estimate. The details of the kernel calculations are the same as in Machens et al. (2001).
(4) Take this estimate as the new stimulus,

$$s_{n+1}(t) = s_n^{\text{est}}(t) \qquad (2)$$

and proceed with step (2).

One hopes that with each iteration the reconstruction error, i.e., the mean-square difference between stimulus and reconstruction, $\langle (s_n^{\text{est}}(t) - s_n(t))^2 \rangle$, becomes smaller. Convergence might be called if the reconstruction error drops below a threshold. A more rigorous criterion is described in the next section. Note that there might exist stimuli that lead to "trivial" solutions. In the absence of spontaneous activity, for instance, a constant subthreshold stimulus $s(t) = s_0$ will fail to elicit any spikes, i.e., $y(t) = 0$. In this case, the parameter choice $h = s_0$ and $k(\tau) = 0$ leads to a perfect fit of the stimulus. This is not

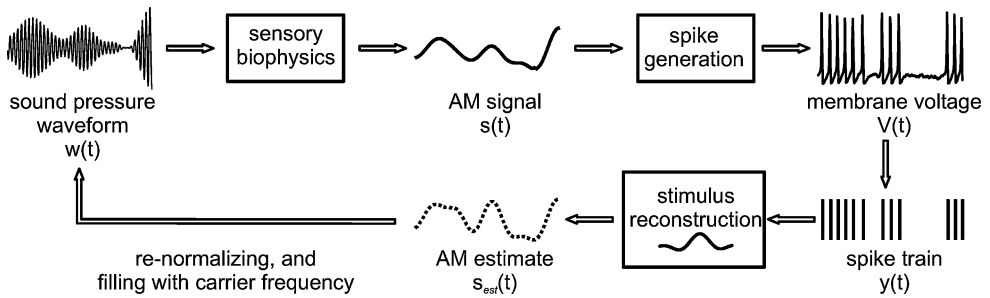

Figure 1. The iterative reconstruction algorithm. The amplitude modulation $s(t)$ of a sound pressure waveform $w(t)$ is extracted by the sensory biophysics. It is this amplitude modulation that we refer to as stimulus. The spike generation mechanism of the receptor neurons transforms $s(t)$ into a voltage signal $V(t)$ which is recorded. The spike train $y(t)$ is extracted online and used to reconstruct the original amplitude modulation. This stimulus estimate $s_{\text{est}}(t)$ is re-normalized, filled with a carrier frequency and played back to the animal after which the procedure is iterated.

a flaw of the algorithm: if the spontaneous activity of the cell is sufficiently low, the "absence" of a stimulus is encoded with high reliability, thus allowing perfect reconstruction. Moreover, the absence of a stimulus is certainly of biological relevance. Nevertheless, the result is trivial in the sense that it does only provide information about the sensory system that could also have been inferred from the neuron's stimulus-response curve.

When starting with stimuli that do excite the neuron, there is still the possibility of converging onto this trivial solution: a non-perfect reconstruction always results in a loss of stimulus variance; hence, the reconstruction algorithm will slowly reduce the stimulus power. To avoid this spurious effect, we readjusted the variance of the stimuli. This is done by modifying step (4):

(4a) Take the renormalized estimate as the new stimulus,

$$s_{n+1}(t) = \frac{\sigma_n}{\sigma_n^{\text{est}}}\left(s_n^{\text{est}}(t) - \mu_n^{\text{est}}\right) + \mu_n^{\text{est}} \qquad (3)$$

where σ_n denotes the standard deviation of the original stimulus $s_n(t)$ while σ_n^{est} and μ_n^{est} denote standard deviation and mean of the reconstructed stimulus. Then proceed with step (2).

To measure the success of each reconstruction, we used the normalized reconstruction error,

$$\varepsilon_n = \frac{\left\langle \left(s_n^{\text{est}}(t) - s_n(t)\right)^2 \right\rangle}{\langle (s_n(t) - \langle s_n(t) \rangle)^2 \rangle}. \qquad (4)$$

Hence, a reconstruction error of $\varepsilon = 0\%$ is equivalent to a perfect reconstruction of the stimulus; a reconstruction error of $\varepsilon = 100\%$ implies that no aspect of the stimulus could be recovered from the spike train.

2.2. *Convergence and Reliability*

Technically, the algorithm can be described by an iterative equation $s_{n+1}(\cdot) = F[s_n(\cdot)]$, where the functional F denotes the process of presenting the stimulus $s_n(t)$, recording the response and reconstructing the stimulus. Starting with an initial stimulus $s_1(t)$, we are searching for a fixed-point solution to this equation which is given if $s_{n+1}(t) \equiv s_n(t)$. Whether such fixed points actually exist is an empirical question, see also Section 2.4.

In the real world, neural systems are contaminated with noise and the iterative equation becomes stochastic. If we assume that the system under investigation is stationary with finite memory, then the individual iterations are fully described by a conditional probability density $p(s_{n+1}(\cdot) \mid s_n(\cdot))$. Hence, the iterative algorithm corresponds to a Markov process and a fixed point is reached if the Markov process settles into a steady state. If we assume that the initial stimulus is drawn from a stationary density $p(s_1(\cdot))$, as is the case for white noise, we can formulate the following necessary and sufficient requirement for convergence:

(5) The iterative algorithm has converged if the stimuli $s_{n+1}(t)$ and $s_n(t)$ have identical statistical properties.

Note that the reconstruction error does not reach zero but instead levels out at a non-zero value that depends on the width of the conditional densities. While convergence of the stimulus statistics ensures convergence of the reconstruction errors, the reverse is not true: even if the reconstruction errors converge to a finite value, the stimulus statistics might still change.

The convergence properties of the algorithm can be demonstrated by simulations with standard integrate-and-fire model neurons (Koch, 1999). Firing threshold was set to 15 mV above the reset potential after spike generation, and an RC-time constant of 10 msec was chosen. Neural refractoriness was described by an absolute refractory period of 3 msec, during which spike generation was turned off. Internal noise sources were incorporated by stochastically varying the firing threshold according to a Gaussian probability distribution whose width is denoted as "internal noise level". For simplicity, the magnitude of external stimuli is always given as a voltage. In some simulations, the input stimuli were convolved with a Gaussian filter with a width of 2.5 msec.

To start the iterative reconstruction algorithm, we chose an initial stimulus that was a white-noise current injection with a cut-off frequency (500 Hz) well beyond the inverse of the time constant of the model neurons. After $n = 100$ iterations, the statistics of the stimuli have converged and the final reconstruction error reflects the noise of the system as shown in Fig. 2a.

Strictly spoken, the assumption of stimulus stationarity only applies for the limit of infinitely long stimuli. For short stimuli, finite-size effects set in as shown in Fig. 2b and c. Here the iterative algorithm was tested on integrate-and-fire neurons using stimuli with different length. As shown exemplary by the reconstruction error, skewness and curtosis of the stimuli after $n = 100$

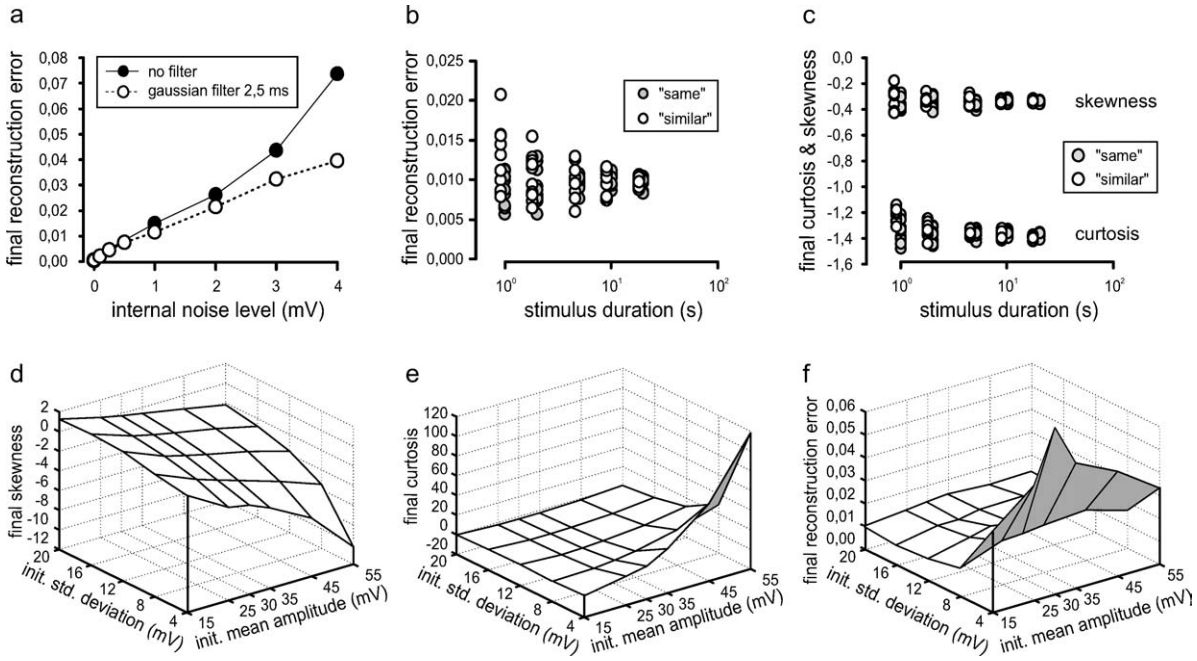

Figure 2. Convergence and reliability of the algorithm. Since neural systems are noisy, the reconstructed stimulus can never be a perfect copy of the original stimulus. Accordingly, the iterative reconstruction procedure levels out at a finite reconstruction error. Using simulations with integrate-and-fire model neurons, panel (a) demonstrates that this final reconstruction error depends on the amount of internal noise in the system. Moreover, for noisy systems, the iterative algorithm cannot converge onto the same stimulus in two independent runs, even if the initial stimuli are identical. It will however, converge onto stimuli that share the same statistical properties. This is also true if the initial stimuli are different but share the same statistical properties. The reliability of this convergence is determined by the length of the stimuli. Panels (b) and (c) demonstrate these observations using integrate-and-fire neurons. Panel (b) depicts the final reconstruction errors, panel (c) shows the higher-order statistics (skew and curtosis) of the amplitude distribution of the final stimuli. In general, there might exist several solutions of the iterative reconstruction algorithm with different overall properties. Panels (d) and (e) show for simulations with a Gaussian input filter (width 2.5 msec) that skew and curtosis of the amplitude distribution of the final stimuli depend on the mean and variance of the initial stimuli. Here, "same" denotes a start with identical stimuli, and "similar" a start with stimuli whose statistical properties are identical. Nevertheless, the final reconstruction errors are similar and small, except for stimuli with very small variance (f). Similar results were obtained for simulations with other model parameters.

iterations, sufficiently long stimuli ($T \geq 10$ sec at a firing rate of at least 30 Hz) result in the same final stimulus statistics with high reliability whereas shorter stimuli produce far more scatter.

For the experiments on auditory receptor neurons, we chose a stimulus length of $T = 10$ sec. For practical reasons, we mostly terminated the iterative reconstruction algorithm after a fixed number of iterations. Some experiments were allowed to run until the recorded neurons were lost. Thus, it was possible to assess the convergence of the algorithm and check how stable the statistical properties of the stimuli became.

2.3. Choice of Initial Stimuli

Ideally, one would start the algorithm close to a steady-state solution, i.e., with a stimulus whose statistics is similar to that of the steady-state stimuli. In the absence of any such knowledge about the system, we suggest to start the iterations with a fairly "general" stimulus such as white noise. Note that the algorithm cannot uncover any stimulus features that are not included in the initial stimulus $s_1(t)$. For instance, if the original stimulus has a cut-off frequency of 50 Hz, then the final stimulus cannot have any frequencies beyond 50 Hz. Hence, the cut-off frequency of the white-noise stimulus should be beyond the inverse of the shortest time constants of the system.

For arbitrary systems, the iterative algorithm might have many possible steady-state solutions. In this case, the choice of the initial stimulus determines into which steady state the iterative algorithm relaxes. For the simulated integrate-and-fire neurons, we investigated the dependence of the final stimulus statistics on the initial

stimulus statistics (see Fig. 2d and e). The data demonstrate that the higher-order statistics of the final stimuli depend non-trivially on the statistics of the initial stimulus. This shows that the optimization landscape has several local minima. As illustrated in Fig. 2f, most of these minima have similarly small reconstruction errors, only for small variance of the stimuli do the reconstruction errors increase.

2.4. Limitations

The iterative reconstruction algorithm developed here is a straightforward extension of linear reconstruction methods. Hence, any limitation that apply to these methods also apply to the new algorithm. For instance, if the system encodes stimulus features that cannot be reconstructed with linear methods, then the outcome of the algorithm is essentially undefined. Hence, there is no prior guarantee that the algorithm will converge as described in point (5) in Section 2.2. Rather, convergence can only be determined empirically.

2.5. Electrophysiology

All experiments were carried out in male and female *Locusta migratoria*, with head, legs and wings cut off, waxed dorsal side up on a temperature-controlled metal platform. To minimize animal movement, the intestines were removed. The frontal upper part of the thorax and fat tissue covering the nervous system were removed to allow for access to the auditory nerve from the dorsal side. The nerve was tethered with a special forceps. Intracellular recordings were performed with standard glass micro electrodes (filled with 1 M KCl solution, impedance 50 to 100 MΩ) on a NPI BRAMP-01 amplifier. The amplifier output was high-pass filtered and fed to the analog inputs of a National Instruments MIO 16-E1-PCI data acquisition board. The whole setup was electrically and acoustically shielded as well as acoustically calibrated (1–40 kHz).

Auditory receptor neurons were identified by determining their frequency sensitivity via a threshold curve. This was measured by a binary search algorithm (3 stimulus repetitions at each step) to find the lowest response intensity for each frequency (1–40 kHz). Recordings were stable up to 90 min.

The stimuli used were 10-second long, random amplitude modulations of a sine-wave carrier whose frequency was given by the best frequency of the cell. The average signal amplitude was set to a certain level above the neuron's measured response threshold at the best frequency. The conversion rate for the signal output was 250 kHz, attenuation of signals was done via a self-built programmable analog attenuator. The laboratory software used (OEL) generated the stimuli for each iteration, controlled the signal output via the analog output ports of the utilized A/D board and the attenuator, recorded the amplifier output voltage and other signals, and automatically detected spikes in the recorded voltage trace by applying a threshold. The recorded spike train was used to compute the reconstruction online as described above. The reconstructed stimulus was then played to the animal. Due to computing requirements, there was a silent pause of approximately 1 second between iterations.

3. Results

3.1. A Sample Reconstruction Series

We tested the iterative reconstruction algorithm on locust auditory receptor neurons. A typical evolution of stimuli and spike trains is shown in Fig. 3 for a sample reconstruction series. The initial stimulus consists of Gaussian white-noise amplitude modulations of a sine-wave carrier. The amplitude modulations have a standard deviation of 7 dB, a mean intensity of 5 dB above the cell's threshold, and a cut-off frequency of 200 Hz, as can be seen by the initial stimulus spectrum, cf. Fig. 3, top row, central panel.

The initial stimulus and the resulting spike train are displayed in the top row, left panel, of Fig. 3. After computing the reconstruction kernel (see inset), an estimate of the original stimulus is obtained as described in the section on materials and methods. The reconstructed stimulus is then normalized and presented as the new stimulus (see also Fig. 1). Although the reconstructed stimulus deviates significantly from the original stimulus (reconstruction error $\varepsilon = 64\%$), the spike train elicited by the reconstructed signal is almost exactly the same as the spike train elicited by the original stimulus, compare the upper two rows of Fig. 3. This result is by no means trivial; instead it serves to show that the reconstructed stimulus itself was a likely candidate for eliciting the original spike train. The similarity of the first two spike trains supports the general validity of the stimulus reconstruction approach.

Remarkably, from each iteration to the next, the spike trains change only slightly. In some cases spikes are

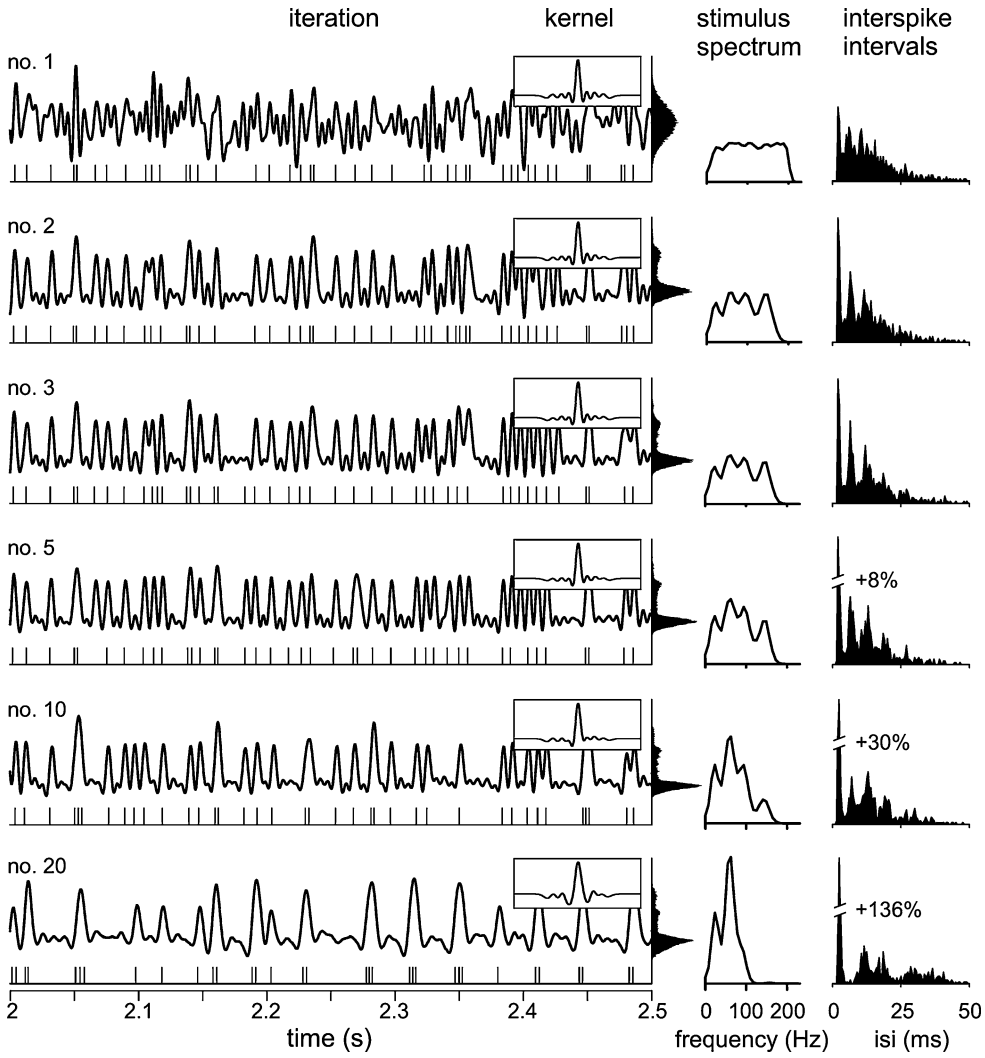

Figure 3. A sample reconstruction series from auditory receptor neurons. Shown are (from top to bottom) the initial stimulus presentation, the subsequent second, third, fifth, and tenth iterations, as well as the final, twentieth stimulus. For each row, the left panel illustrates the stimulus $s(t)$ together with the spike train it evoked, and (in the inset) the reconstruction kernel calculated from this stimulus and spike train. Directly attached to the right is the distribution of modulation amplitudes. The central panel depicts the spectrum of the stimulus and the right panel the distribution of interspike intervals. During the iterations, the stimulus loses high-frequency content and develops a strongly non-Gaussian amplitude distribution, leading to long sub-threshold parts interrupted by brief sound pulses. The interspike-interval histogram develops a strong peak around 2.5 ms, corresponding to a burst-like firing pattern. To resolve the full structure of the interspike-interval histograms, the first peak is not drawn to scale. The relative fraction cut out of the peak is indicated by the percentage number.

shifted, usually for a fraction of a millisecond only. At other instances spikes disappear or new ones are inserted. In the long term, these small alterations eliminate spike patterns that are not stable to perturbations by internal neural noise sources or that are less well reconstructed than others. In particular, the most unstable patterns will vanish within the first iterations; more stable patterns might exist for a longer time, while only the most stable are likely to remain at all. As can be seen in Fig. 3, it turns out that patterns of two or more spikes are more stable than single spikes. When comparing iteration #10 with iteration #20, for instance, one sees that all double- and triple-spike patterns of iteration #10 persist to #20, yet almost none of the single-spike patterns does. Hence, after about 20 iterations, only burst-like patterns remain.

These long-term changes are reflected in corresponding changes of the stimulus and spike-train statistics. Initially, the stimulus is Gaussian distributed and its power is shared evenly among all frequencies between 0–200 Hz (cf. Fig. 3, central panel, top row). During the iterated stimulus reconstructions, the high frequencies are slowly depleted, leaving the final stimulus with frequencies between 0–100 Hz only. The amplitude distribution develops a strongly non-Gaussian shape with a pronounced peak at lower, sub-threshold intensities. Concurrently, the spike train changes from an initially Poisson-like process into one that is dominated by groups of spikes. This trend is clearly visible in the interspike-interval distribution as shown in Fig. 3 on the right. An initially almost exponential distribution evolves into one with a high peak at 2.5 ms, corresponding to burst-like firing, and fewer long interspike intervals.

Figure 4 illustrates the development of the reconstruction error as well as the average firing rate for the sample reconstruction series. After a strong initial drop, the reconstruction error decreases only slightly from $\varepsilon \sim 20\%$ (iteration #2) to $\varepsilon \sim 10\%$ (iteration #20). Hence, most of the reconstruction success can be attributed to eliminating stimulus features not encoded by the system (drop from $\varepsilon = 64\%$ to $\varepsilon = 20\%$ during the first iteration), and only a comparably smaller gain is achieved by eliminating stimulus features less robust against noise (decrease from $\varepsilon = 20\%$ to $\varepsilon = 10\%$ during 19 further iterations).

For this example, the firing rate is approximately constant. This shows that the improved reconstructability of the stimulus is not the result of an increase of the average firing rate, but rather the consequence of a decrease in stimulus bandwidth and an increase of periods where the stimulus is below threshold.

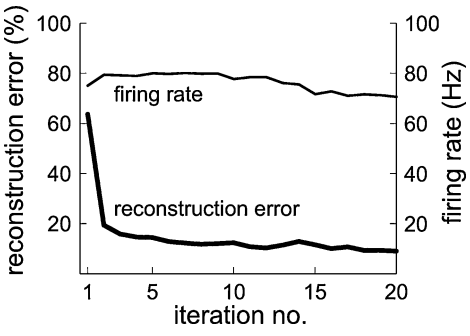

Figure 4. Development of the reconstruction error and firing rate for the experiment shown in Fig. 3. The largest decrease of the reconstruction error occurred immediately after the first iteration; subsequent iterations lead only to a slight further decrease. In this experiment, the firing rate remained approximately constant.

3.2. Population Data and Convergence Properties

Altogether 14 cells were recorded. In most experiments, 20 iterations of the procedure were carried out; longer sequences (up to 71 iterations) were used to evaluate the long-term convergence of the iterations. While all initial stimuli had a cut-off frequency of 200 Hz, both the mean above threshold and the overall variance were varied to analyze the influence of the initial conditions on convergence.

In all cases the reconstruction errors decrease significantly. Pooled over all experiments with more than ten iterations ($n = 18$), the reconstruction error of the initial stimulus is $\varepsilon = 62.0 \pm 13.4\%$ and of the last stimulus $\varepsilon = 9.4 \pm 3.0\%$. In most experiments, the firing rate stays either relatively constant throughout the iterations or decreases slightly. While initially there is a strong correlation between high firing rates and low reconstruction errors, the final reconstruction errors are largely independent of the firing rate. In some experiments, the reconstruction error did at times increase between iterations (cf. also Fig. 4).

From the results presented up to now, one might infer that after a sufficient number of iterations, a stimulus/response pattern evolves that is unique and characteristic for the neuron under study. However, this is not the case. The types of stimuli that emerge rather depend on the initial condition chosen. The final stimuli fall into two main classes (Fig. 5a and b).

For initial conditions in which more than 25% of the amplitude modulation values are below the cell's threshold (as in the example shown in Fig. 3), a stimulus emerges in which the evoked spikes gather in groups of three to five spikes with interspike intervals given by the neuron's refractory period ($\approx$2.5 ms). The final stimulus consists of many sub-threshold periods, interrupted by brief yet strong pulses of highly uniform amplitude and shape (cf. Fig. 5a). Furthermore, the stimulus spectra typically contain broad-band frequencies in the range 0–100 Hz. Within this group, the reconstruction error is $\varepsilon = 9.7 \pm 2.7\%$.

Initial stimuli with almost no amplitude modulation values below threshold ($<5\%$) lead in most cases to a second type of final stimulus (Fig. 5b) which consists of slow, regular waves in the amplitude modulation of the stimuli, with a narrow-banded spectrum centered

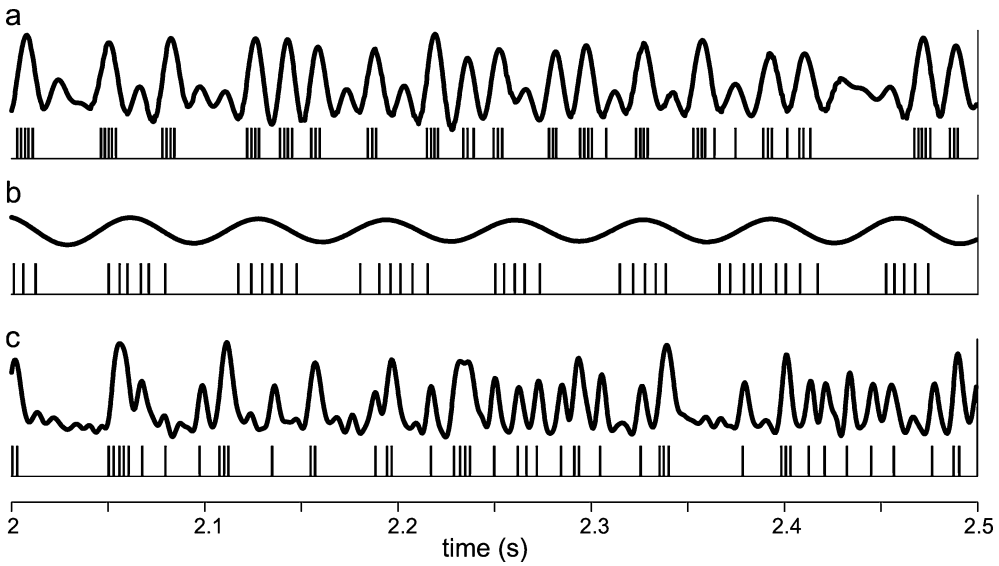

Figure 5. Final Stimuli. Shown are the stimulus and spike train of the final iteration for three exemplary experiments. Final stimuli consist either of short bursts (a) or of oscillatory activity (b). Some of the experiments also end with stimuli as shown in (c); however, these are likely to have not yet converged.

around 20–50 Hz. The resulting long-range correlations of the stimulus are reflected in the oscillatory lobes of the reconstruction kernel. The evoked spikes occur in long groups with a frequency that depends on the actual wave amplitude. The final reconstruction errors are slightly lower ($\varepsilon = 7.8 \pm 3.9\%$) and can be as low as $\varepsilon = 2\%$.

In some of the experiments, the final stimuli belonged to none of these classes. These stimuli consist of pulses of various amplitudes and duration that evoke between one and five spikes each (Fig. 5c). Hence, in contrast to the first stimulus type, these stimuli still elicit many solitary spikes. Moreover, the stimuli retain higher frequency components up to 160 Hz. However, careful inspection of the traces shows that similar stimuli do appear frequently in intermediate stages of all reconstruction series. Hence, we conclude that in these experiments, the reconstruction series may not yet have converged to one of the previous types. Note that sometimes the changes from one stimulus to the next were excessively small, thus requiring many more than 20 iterations to converge.

3.3. Simulations with Integrate-and-Fire Model Neurons

In a final step, the iterative reconstruction series was tested numerically on several integrate-and-fire neurons (see Fig. 2). In this case, the initial stimulus was a 10-sec long current injection with a flat spectrum from 0 to 500 Hz. The reconstruction series obtained in these simulations have properties similar to those found in the experimental data. In particular, the reconstruction errors decrease on average, with the biggest jump occuring between the first and second iteration. Moreover, the final stimuli found in the simulations were similar to the ones found for the auditory receptor neurons: while the quantitative details, such as final bandwidth, vary with the model parameters, the prominent, almost binary switching between sub- and supra-threshold stimuli (as witnessed in Fig. 5a) was a common pattern. The reconstruction for the integrate-and-fire neurons also showed a similar dependence on the initial stimuli (Fig. 2d and e) as the auditory receptor neurons.

4. Discussion

Previous online algorithms in neurophysiology have sought to find stimuli that maximally drive a given cell (Harth and Tzanakou, 1974; Nelken et al., 1994). Here we have presented an alternative approach that seeks stimuli that can be decoded with high fidelity. Starting with a general white-noise stimulus, the iterative algorithm eliminates all stimulus features that

cannot be reconstructed reliably. That includes stimulus features not encoded by the particular neuron, stimulus features that are encoded unreliably, and stimulus features that a linear reconstruction algorithm simply fails to decode although they have been encoded. Given our previous work on auditory receptor neurons (Machens et al., 2001), we are confident that the latter type of error can be neglected in the present context.

Our results show that the algorithm is indeed capable of increasing the average reconstruction success with each iteration. However, the lowest reconstruction errors found in the neurophysiological experiments (down to $\varepsilon = 2\%$) could only be observed for low-bandwidth stimuli whose long-range correlations could be exploited by the reconstruction method. More complex stimuli with higher bandwidth did not evolve significantly towards stimuli with errors below $\varepsilon = 10\%$. We conclude that this reflects a residual noise level that cannot be overcome; hence, a perfect reconstruction of "non-trivial" stimuli does not seem to be possible. This conclusion is supported by our numerical studies on integrate-and-fire neurons where we find a dependence of the final reconstruction error on the neuron's noise level (Fig. 2a).

In agreement with earlier findings, our results indicate that higher stimulus frequencies are only stable if the stimulus has a higher variance, leading to larger fluctuations of the amplitude modulations. In the process of the iterative reconstructions, these high-variance stimuli develop long sub-threshold periods interrupted by brief yet forceful sounds. The sound bursts, in turn, result in small groups of rapidly fired spikes. Although even single spikes show a remarkable persistence against perturbations, these small groups of spikes turn out to be the most stable patterns of neural activity.

As most receptor neurons did not show significant spontaneous activity, pauses can be reconstructed with high reliability. One might therefore re-interpret the neuron's performance by saying that the neuron seeks to optimize the duration of potential pauses under the given constraints. Additionally, the spiking behavior directly after the onset of a stimulus is highly reliable. This reliability extends over maximally five spikes (around 12.5 ms) as larger groups of spikes occur only rarely.

The similarity of the final stimuli for both the numerical simulations and the auditory receptor neurons suggest that these stimuli are well reconstructable for a large class of neurons. Interestingly, grasshopper songs have similar properties as they consist of a rapid succession of sound (called syllables) and pauses. Behavioral experiments (Balakrishnan, 2001) have shown that songs with pronounced syllable-pause structure and strong accentuation of the syllable onset (lasting for 15 ms) trigger responses with high reliability. Indeed, grasshopper songs can often be reconstructed with high fidelity (Machens et al., 2001).

The method presented here could also be used as a general exploratory tool to find the most effective electrical stimuli for cells deep in a neural system. Moreover, the method will be helpful to test the performance of stimulus reconstruction methods: if a major fraction of information conveyed by the spike train is not employed by a reconstruction method, then the stimulus will rapidly lose valuable features and degenerate into a trivial solution.

Acknowledgments

We are grateful to Jan Benda for his indispensable help in setting up the software necessary to reconstruct stimuli online. We also thank an anonymous referee for pointing out that the algorithm corresponds to a fixed-point search.

References

Balakrishnan R, von Helversen D, von Helversen O (2001) Song pattern recognition in the grasshopper Chorthippus biguttulus: The mechanism of syllable onset and offset detection. J. Comp. Physiol. A 187(4): 255–264.

Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. Science 252: 1854–1857.

Gabbiani F, Metzner W, Wessel R, Koch C (1996) From stimulus encoding to feature extraction in weakly electric fish. Nature 384: 564–567.

Harth E, Tzanakou E (1974) Alopex: A stochastic method for determining visual receptive fields. Vision Res. 14: 1475–1482.

Koch C (1999) Biophysics of Computation. Oxford University Press, Oxford, UK.

Machens CK, Stemmler MB, Prinz P, Krahe R, Ronacher B, Herz AVM (2001) Representation of acoustic communication signals by insect auditory receptor neurons. J. Neurosci. 21: 3215–3227.

Nelken I, Prut Y, Vaadia E, Abeles M (1994) In search for the best stimulus: An optimization procedure for finding efficient stimuli in the cat auditory cortex. Hearing Res. 72: 237–253.

Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc. R. Soc. Lond. B 262: 259–265.

Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) Spikes—Exploring the Neural Code. MIT Press, Cambridge, MA.

Römer H (1976) Die Informationsverarbeitung tympanaler Rezeptorelemente von Locusta migratoria (Acrididae, Orthoptera). J. Comp. Physiol. 109: 101–122.

Stanley GB, Li FF, Dan Y (1999) Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. J. Neurosci. 19(18): 8036–8042.

Theunissen F, Roddey JC, Stufflebeam S, Clague H, Miller JP (1996) Information theoretic analysis of dynamical encoding by four identified primary sensory interneurons in the cricket cercal system. J. Neurophys. 75(4): 1345–1364.

Wessel R, Koch C, Gabbiani F (1996) Coding of time-varying electric field amplitude modulations in a wave-type electric fish. J. Neurophysiol. 75(6): 2280–2293.