LETTER Communicated by Maneesh Sahani

Abstract Stimulus-Specific Adaptation Models

Robert Mill
robert.mill@plymouth.ac.uk
School of Psychology and Centre for Robotics and Neural Systems, University of Plymouth, Plymouth PL4 8AA, U.K.

Martin Coath
martin.coath@plymouth.ac.uk
Thomas Wennekers
thomas.wennekers@plymouth.ac.uk
School of Computing and Mathematics and Centre for Robotics and Neural Systems, University of Plymouth, Plymouth PL4 8AA, U.K.

Susan L. Denham
susan.denham@plymouth.ac.uk
School of Psychology and Centre for Robotics and Neural Systems, University of Plymouth, Plymouth PL4 8AA, U.K.

Many neurons that initially respond to a stimulus stop responding if the stimulus is presented repeatedly but recover their response if a different stimulus is presented. This phenomenon is referred to as stimulus-specific adaptation (SSA). SSA has been investigated extensively using oddball experiments, which measure the responses of a neuron to sequences of stimuli. Neurons that exhibit SSA respond less vigorously to common stimuli, and the metric typically used to quantify this difference is the SSA index (SI). This article presents the first detailed analysis of the SI metric by examining the question: How should a system (e.g., a neuron) respond to stochastic input if it is to maximize the SI of its output? Questions like this one are particularly relevant to those wishing to construct computational models of SSA. If an artificial neural network receives stimulus information at a particular rate and must respond within a fixed time, what is the highest SI one can reasonably expect? We demonstrate that the optimum, average SI is constrained by the information in the input source, the length and encoding of the memory, and the assumptions concerning how the task is decomposed.

1 Introduction

Much research over the past half-century has been directed toward understanding how single biological neurons process information, both with and

without the context of a network (Koch, 1998). One widespread property of neurons, which is increasingly gaining recognition as an implicit form of computation, is stimulus-specific adaptation (SSA). SSA refers to a reduction, or total cessation, in the spiking response of a single neuron to a repeated stimulus, which recovers in response to the presentation of another, rare stimulus (Ulanovsky, Las, & Nelken, 2003). The SSA response of single neurons implicitly manifests a form of memory capable of storing at least the statistics of recent stimuli. SSA may therefore represent a neural correlate of other aspects of novelty detection, such as the mismatch negativity (MMN) component seen in event-related potential signals in response to oddball stimuli (Nelken & Ulanovsky, 2007) (but see Winkler, Denham, & Nelken, 2009), or the priming effect in behavioral psychology (Grill-Spector, Henson, & Martin, 2006).

In SSA experiments, one is concerned to infer from the spikes measured in a neuron not the identity of the stimulus as such—in fact, it may be the case that the response of the neuron carries no information about the identity of a stimulus—but rather the rarity of the stimulus in relation to a background of recent stimuli. A single-valued measure of the selectivity that a neuron exhibits as it adapts to stimuli presented in an oddball sequence is the stimulus-specific adaptation index (SI) (see section 2.1, equation 2.1), which was originally proposed by Ulanovsky et al. (2003); Ulanovsky, Las, Farkas, & Nelken (2004). Although the original choice of function was presumably intuitive and somewhat ad hoc, it has subsequently been adopted by other researchers (Von der Behrens, Bäuerle, Kössl, & Gaese, 2009; Anderson, Christianson, & Linden, 2009; Malmierca, Cristaudo, Pérez-González, & Covey, 2009) and has been applied to measurements of evoked local field potentials as well as single neuron responses (Von der Behrens, Bäuerle, Kössl, & Gaese, 2009).

The SI value can be viewed as a measure of how sensitive the response of a neuron is to the statistical distribution of stimuli in a sequence: a large SI value indicates that a neuron responds more vigorously to rarer stimuli. Neurons that respond to novel events in the environment are likely to confer an evolutionary advantage on an organism, enabling it to both respond to sudden changes that could indicate a predator or prey and conserve energy through desensitization to common events. In this letter, we broaden this view to encompass any kind of computational process that attempts to identify a stream of input symbols one by one, store them in a memory, and assign a higher output value to rare symbols than to common ones. A computation of this kind we refer to as an abstract SSA model (ASSAM)—a theoretical "machine" that emits spikes in response to an input pattern in a way that maximizes the SI on average. As the biological mechanism of SSA is still poorly understood, referring in the interim to an abstract, rather than a concrete, process permits one to address questions within the traditional frameworks of probability and information theory, such as, "If incoming symbols are frequently misclassified, or the content of the memory is faulty,

what is the impact on the SI?" and "How does the length of the memory affect the SI?" and so on. The motivation behind this approach stems from a need to identify exactly what it is that contributes to—and limits—the SI value. In an earlier study (not reported), SI values were computed for units in a model of SSA that took the form of a network of integrate-and-fire neurons. When only low SI values could be obtained, the question arose concerning the extent to which the SI values depended on the configuration of model on the one hand and the information content of the input on the other.

We consider that this work contributes to two areas of research. The first is practical and was alluded to above. Large-scale computational models of neural systems can take a long time to design and execute. If it can be shown that for a given input, the SI achieved by the ideal system (i.e., an ASSAM) cannot exceed a certain limit on average, then one can be assured that no other system will either, regardless of its implementation. The second contribution lies in making explicit the precise nature of the computation that a neuron is said to perform when it exhibits enhanced responses to rare stimuli. Others have undertaken the task of quantifying formally the information that must be present in a neural input signal for a system to discriminate stimuli with a given effectiveness (Seung, & Sompolinsky, 1993; Brunel, & Nadal, 1998; see Dayan, & Abbott, 2001, for a review). This letter reports a similar investigation into the ability of a system to respond to novel events, with special regard to an experimental format (the oddball paradigm; see section 2.1) and summary statistic (the SI metric) that are popular in the physiological research community.

The organization of the letter is as follows. Section 2 provides a brief review of the stimulus-specific adaptation studies that have been conducted to date with a particular focus on those that make use of the oddball paradigm, as it is in this context that the SI value finds its most natural definition. Section 3 provides a formal definition of an abstract SSA model. Section 4 explains how SI values can be predicted for ASSAMs that are supplied only with information concerning which of two neurons (modeled as counters) is the first to reach threshold and fire. This model is examined for three types of memory: in the first, it is informed a priori which stimulus is the deviant; in the second, it infers what the standard (and, by implication, the deviant) is, based on which stimulus appears most frequently in a buffer of recent stimuli; and in the third, the memory is a dynamical system that undergoes phases of adaptation and recovery designed to mimic two depressing synapses. We show how the discriminability of the input and the size and type of memory conspire to place an upper limit on the expected SI value that the ASSAM—and hence any other algorithm based on the same principles—can achieve. Section 5 extends the results of section 4 to accommodate models that receive more than two Poisson inputs. Section 6 discusses the explicit maximization of the SI with respect to input patterns. Finally, in section 7,

experimental results are reconsidered in the light of these theoretical insights.

2 Stimulus-Specific Adaptation

Adaptation following the first few presentations of a stimulus is ubiquitous in biological systems and manifests itself on various scales. The term *habituation* is used to refer to a decrement in the behavioral response to a stimulus after a sustained presentation, and *dishabituation* refers to a recovery in the response to a new stimulus (Nicholls, Martin, Wallace, & Fuchs, 2001). Thus, behavioral habituation and dishabituation are characteristically similar to SSA observed in single neurons. The response of neurons in the auditory pathway to repetitive tone stimuli has been the subject of particularly intensive research.[1] For instance, in all subdivisions of the inferior colliculus (IC), one finds cells that cease responding to a train of identical tones if the tones are presented at an appropriate rate. Furthermore, a subset of these cells shows a recovery in their response if a different tone is presented, which is referred to as a *novelty response*, and adaptation is then said to be stimulus-specific (Pérez-González, Malmierca, & Covey, 2005).

SSA responses have also been discovered in auditory thalamus (Anderson et al., 2009) and auditory cortex (Von der Behrens et al., 2009; Ulanovsky et al., 2004, 2003). The mechanism underlying SSA is unknown. Firing-rate adaptation has been widely dismissed as a candidate due to its inability to account for stimulus specificity (e.g., Grill-Spector et al., 2006; Ulanovsky et al., 2004). Synaptic depression may be implicated in SSA, especially if it can be shown that the signals relating to different stimuli are directed along distinct synaptic routes (see the discussion in section 7).

2.1 Oddball Experiments and the SSA Index. One consequence of SSA is that the response of a neuron to a tone sequence reflects the statistics of the sequence. For example, if tone A appears often in a sequence of tones and tone B appears rarely, then, on average, the neuron will typically adapt to A more than it does to B, and hence will develop a preference for tone B. In order to ensure that the preference is for the novelty of the tone, rather than the tone itself, one compares the spike counts elicited by the same tone in different contexts using the novelty response (Pérez-González et al., 2005) or oddball (Ulanovsky et al., 2003) experimental paradigms.

In an oddball experiment, two sequences consisting of tones A and B are presented, and the spike counts in response to each tone are recorded for a single neuron (see Figure 1). In the first sequence, tone A appears frequently (standard), and tone B appears rarely (deviant); in the second sequence, the

[1] As auditory stimuli are typically in view throughout this article, we will typically use the terms *stimulus* and *tone* interchangeably.

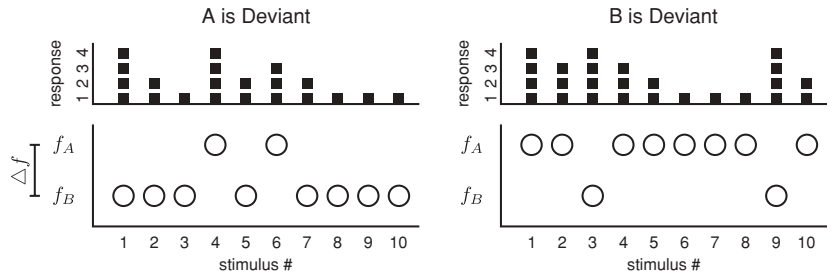

Figure 1: Illustration of an oddball experiment, in which two sequences of 10 tones are presented. In the first sequence, tone A is deviant; in the second, tone B is deviant. The lower panels depict the order of presentation; the upper panels plot the number of spikes elicited. In this example, $d_A = 3.50$, $s_A = 2.13$, $d_B = 4.00$, $s_B = 1.75$, and SI $= 0.32$ (2 decimal places).

probabilities are reversed, and tone A is deviant and tone B is standard. (A third, control block is sometimes presented in addition, in which tones A and B appear with equal probability, but we do not need to consider the control block in this work.)

Having measured the spike counts for each tone in an oddball experiment, the next step is to use these values to determine the sensitivity of the neuron to the rarity of a tone. Ulanovsky et al. (2003) have defined the stimulus-specific adaptation index (SI) as

$$\mathrm{SI} = \frac{d_A - s_A + d_B - s_B}{d_A + s_A + d_B + s_B}, \tag{2.1}$$

where d_x and s_x, $x \in \{A, B\}$, are the mean spike counts elicited per tone in response to frequency f_x when presented as deviant and standard, respectively. SI is restricted to $-1 \leq \mathrm{SI} \leq 1$, where $\mathrm{SI} > 0$ and $\mathrm{SI} < 0$, respectively, indicate sensitivity to a deviant or a standard. As special cases, $\mathrm{SI} = 0$ for a neuron that responds preferentially to neither standard nor deviant tones and $\mathrm{SI} = \pm 1$ for a neuron that responds exclusively to deviants or standards, respectively. Figure 1 provides an example of two oddball sequences, in which one in five tones is deviant (i.e., the probablity of a deviant, p_{dev}, is 0.2). The SI and its components are listed in the caption.

2.2 Principles Established by SSA Studies. SSA studies typically focus on how the SI value is affected by changes to the oddball stimuli parameters, such as the probability of a deviant (p_{dev}), the frequency difference between the tones ($\triangle f$, octaves), and the tone rate. The findings from the experimental studies considered (Von der Behrens et al., 2009; Anderson et al., 2009; Malmierca et al., 2009; Ulanovsky et al., 2004, 2003) vary in

respect to their details but nevertheless appear to reveal a set of common relations between these basic parameters and the SI.

- *Discriminability principle.* The SI value is larger when the frequency separation between tones in the oddball sequence, $\triangle f$, is larger. When $\triangle f = 0$, the tones are indistinguishable, and SI $= 0$.
- *Rarity principle.* For the experimental parameters investigated in the literature, the SI increases as the deviant tones become rarer, that is, as p_{dev} decreases toward zero. It is not clear how the SI continues to increase as p_{dev} approaches (without reaching) zero; however, the limiting behavior in the other direction is clear: if $p_{dev} = 0.5$, neither tone is deviant, and SI $= 0$.
- *Locality principle.* By definition, any SSA process requires a memory; however, experimental evidence also suggests that for biological neurons, this memory is finite. Ulanovsky et al. (2004) have performed an experiment in which the deviant and standard tones are swapped at regular intervals (the switching blocks experiment). After exposure to a short duration of the new stimulus statistics, enhanced spiking is observed in response to the new deviant, indicating that the most local tones are used to establish the identity of the deviant. If the statistics of the entire sequence were accounted for when responding to each tone, then the inferred, global p_{dev} would gradually tend toward 0.5 and the SI value toward zero.

2.3 The Expected SSA Index. The SI described above is a function that quantifies the degree to which a neuron exhibits SSA on the basis of four fixed measurements. However, as the number of spikes counted in response to a tone is random and thus varies from trial to trial, it is appropriate to replace the measurements, d_A, s_A, ..., with random variables, D_A, S_A, ..., and consider the properties of the derived, random variable, SI.

It is important at this juncture to make explicit the dependence of all these quantities on the total number of tones presented, n, and so instead write SI_{n}, $D_{A,\mathsf{n}}$, ... Then one can write the expected value of the SSA index for n tones as

$$E\{SI_{\mathsf{n}}\} = E\left\{\frac{D_{A,\mathsf{n}} - S_{A,\mathsf{n}} + D_{B,\mathsf{n}} - S_{B,\mathsf{n}}}{D_{A,\mathsf{n}} + S_{A,\mathsf{n}} + D_{B,\mathsf{n}} + S_{B,\mathsf{n}}}\right\}. \tag{2.2}$$

Note that $E\{D_{A,\mathsf{n}}\}$ is the expected sample mean spike count per tone over the first n tones for deviant A, with $E\{S_{A,\mathsf{n}}\}$, $E\{D_{B,\mathsf{n}}\}$ and $E\{S_{B,\mathsf{n}}\}$ similarly defined.

The SSA experiments reported in the literature tacitly assume that n is so large that using a larger value would not significantly influence the mean of the SI. Following this principle, we define the expected SI (ESI) as the value to which one expects the SI to converge over an oddball sequence of

infinite length, that is,

$$\begin{aligned}\mathrm{ESI} &\triangleq E\{SI_\infty\} \\ &= E\left\{\frac{D_{A,\infty} - S_{A,\infty} + D_{B,\infty} - S_{B,\infty}}{D_{A,\infty} + S_{A,\infty} + D_{B,\infty} + S_{B,\infty}}\right\} \\ &= \frac{E\{D_A\} - E\{S_A\} + E\{D_B\} - E\{S_B\}}{E\{D_A\} + E\{S_A\} + E\{D_B\} + E\{S_B\}}. \end{aligned} \tag{2.3}$$

Recall that the random variables $D_{A,\mathsf{n}}$, $S_{A,\mathsf{n}}$, ..., are sample means obtained for n samples, so their variances tend to zero as $\mathsf{n} \to \infty$. The expected SI is therefore a function of expected spike counts, $E\{D_A\}$, $E\{S_A\}$, ..., for which one can obtain analytical expressions in the case of simple models. A formal framework for describing these models is presented next.

3 Abstract SSA Models

There are various ways that SSA could be investigated using computational models. For instance, to achieve a direct analog of the physiological experiments described earlier, one might simulate the response of a population of cells with various synaptic connectivities to oddball stimuli and then search for individual units in the network whose SI value significantly exceeds zero. One could then go on to investigate how the SI is affected by changing the frequency separation between the tones or the probability of a deviant. In what follows, we pursue the alternative approach of viewing SSA as a task to be performed—whether by a digital computer, a network of biological cells, or with pen and paper—and consider what procedure maximizes the ESI for a given input.

On reflection, it is first of all clear that in order to perform this task, it must be possible to discriminate tones on the basis of some kind of evidence. In physiological terms, this evidence is restricted to the pattern of neural impulses initially encoded at the auditory periphery. In this letter, we consider a scheme in which tones are encoded in the firing rates of Poisson processes (see section 4). The encoding process is taken to be noisy, such that the relationship between an input tone, $X \in \{A, B\}$, and the evidence supplied to the SSA process, $\epsilon \in \mathcal{E}$, is best characterized using conditional probability functions: $p(\epsilon \mid X = A)$ and $p(\epsilon \mid X = B)$.

Second, without the ability to recall the evidence recorded in the recent past, it is impossible to decide whether the tone just presented was deviant. Thus, we also endow the model with a memory, letting $\mathcal{M}$ denote the set of all possible states that the memory can occupy for a given memory encoding scheme (e.g., the state of all synapses in a model).

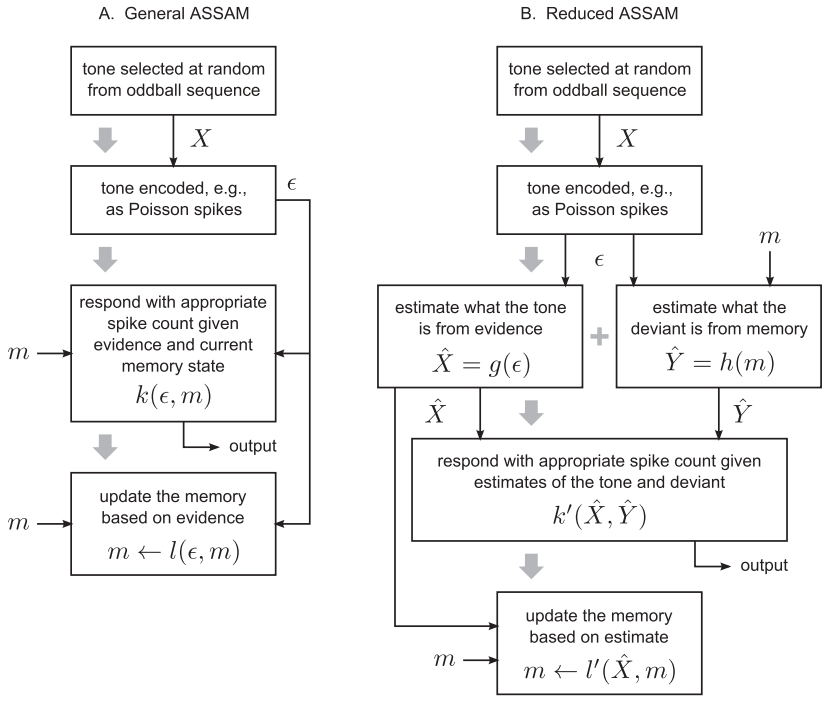

Figure 2: Flow diagrams for the (A) general and (B) reduced abstract SSA models. The solid gray arrows depict the order in which processing stages are carried out, while the thin arcs directed into, and emerging from, boxes indicate the flow of inputs and outputs. Note that the input m refers to the memory state carried over from the previous iteration. The symbols $\hat{X}$ and $\hat{y}$ denote estimations concerning the incoming tone and deviant tone, respectively (see section 3.3).

3.1 General Model. The general ASSAM is a tuple $\mathcal{S} = (m, k, l)$ consisting of a memory $m \in \mathcal{M}$ and two functions, k and l. The function $k : \mathcal{E} \times \mathcal{M} \to \mathbb{N}_0$ maps some input $\epsilon \in \mathcal{E}$ along with a memory state to number of spikes. The function $l : \mathcal{E} \times \mathcal{M} \to \mathcal{M}$ combines arriving evidence with the current memory to generate a new memory state. These two functions in tandem consitute a processing loop that responds to input on the basis of the memory and updates the memory on the basis of the input (see Figure 2A). For this definition to hold, the precise form that $\mathcal{E}$ and $\mathcal{M}$ take need not be specified.

3.2 Reduced Model. The reduced ASSAM is a tuple $\mathcal{S}' = (m, g, h, k', l')$ consisting of a memory $m \in \mathcal{M}$ and four functions. The function

$g : \mathcal{E} \to \{A, B\}$ decides whether the incoming tone was A or B on the basis of evidence $\epsilon \in \mathcal{E}$. The function $h : \mathcal{M} \to \{A, B\}$ decides whether the current deviant is A or B on the basis of the memory m. The function $k' : \{A, B\} \times \{A, B\} \to \mathbb{N}_0$ decides how many spikes to emit on the basis of estimates regarding the incoming tone and deviant. The function $l' : \{A, B\} \times \mathcal{M} \to \mathcal{M}$ updates the memory according to the estimate of the incoming tone (see Figure 2B).

The general ASSAM decides optimally how many spikes to emit on the basis of ϵ and m directly. The reduced ASSAM first decides what the input tone is on the basis of ϵ, then decides what the deviant tone is on the basis of m, and then, on the basis of the outcome of the first two decisions, makes a third, optimal decision concerning how many spikes to emit. This collapsing of evidence into simple decision variables, $\hat{X}$ and $\hat{y}$, reduces the ESI but makes the resulting mathematical descriptions more tractable; consequently, the majority of the ASSAMs considered in this letter are of the reduced kind.

3.3 Relation to Component Probabilities. We demonstrated in section 2.3 that the ESI is a function of the expected number of spikes elicited in response to tone A or B when standard or deviant. For a reduced ASSAM, the expected number of spikes is related to more elementary probabilities.

We use X to denote the identity of the incoming tone. Thus, $X = A$ and $X = B$ denote random events corresponding to the presentation of tone A or B, respectively, so that $P(X = A) + P(X = B) = 1$. Similarly, let $\hat{X} = A$ and $\hat{X} = B$ denote the event that the receiver concludes that tone A or B was presented; thus, we also have $P(\hat{X} = A) + P(\hat{X} = B) = 1$, because the receiver is forced to decide. (We consider here the case of two frequencies. Systems that process multiple frequencies are studied in section 5.) If $\forall x \in \{A, B\}$, $P(\hat{X} = x \mid X = x) = 1$, then the receiver is never mistaken about incoming tones. Such a receiver we refer to as having an ideal discrimination.

Furthermore, we use Y to denote the identity of the tone that is deviant over the course of the entire oddball sequence. For example, $Y = A$ signifies that tone A is the deviant. The p_{dev} parameter used in oddball experiments can then be stated in terms of joint probabilities,

$$p_{dev} = P(X = A, Y = A) + P(X = B, Y = B). \tag{3.1}$$

Similarly, let $\hat{y} = A$ and $\hat{y} = B$ denote that event that the receiver respectively holds tone A or B to be the deviant. If $\forall y \in \{A, B\}$, $P(\hat{y} = y \mid Y = y) = 1$, then the receiver is never mistaken about which tone is the deviant. Such a receiver we refer to as having an ideal memory.

Finally, following the reduced ASSAM scheme, let $k'(\hat{x}, \hat{y})$ denote the number of spikes that the receiver emits when it identifies an incoming

tone as $\hat{x} \in \{A, B\}$ and holds the deviant to be $\hat{y} \in \{A, B\}$. The expected number of spikes emitted when $x \in \{A, B\}$ is presented as a deviant is then

$$E\{D_x\} = \sum_{\hat{x},\hat{y}\in\{A,B\}} k'(\hat{x}, \hat{y}) P(\hat{X} = \hat{x} \mid X = x) P(\hat{y} = \hat{y} \mid Y = x). \tag{3.2}$$

This accounts for all the spikes emitted correctly and in error. Similarly, the expected number of spikes emitted when $x \in \{A, B\}$ is presented as a standard is

$$E\{S_x\} = \sum_{\hat{x},\hat{y}\in\{A,B\}} k'(\hat{x}, \hat{y}) P(\hat{X} = \hat{x} \mid X = x) P(\hat{y} = \hat{y} \mid Y \neq x). \tag{3.3}$$

The expressions for $E\{D_x\}$ and $E\{S_x\}$ in equations 3.2 and 3.3 can be substituted into equation 2.3 to obtain an expression for ESI in terms of stimulus and decision probabilities.

4 SSA with Two Poisson Inputs

The first circuit we consider consists of two Poisson processes, each of which supplies its spikes to a counter-unit, which is equivalent to a non-conductance-based, nonleaky integrate-and-fire neuron. The first counter to receive N input spikes fires a single output spike, and the information concerning which counter fired first is communicated as evidence, $\epsilon \in \mathcal{E} = \{1, 2\}$, to an abstract model of SSA. Readout mechanisms in which two or more hypotheses compete via accumulating evidence to reach a threshold have been proposed previously. For example, Mazurek, Roitman, Ditterich, and Shadlen (2003) have presented a model in which the buildup in support for each hypothesis is tracked in a neural ensemble (as opposed to in the membrane potential of a single neuron). A circuit that estimates the stimulus according to which neuron in a population reaches threshold and fires first has also been referred to as a temporal winner-take-all network (Shamir, 2009) and may be viewed as a special case of a rate-based winner-take-all network (Arbib, 1995), in which the excitatory and inhibitory feedback connections are both rapid and efficacious (see section 5.2).

The rate of Poisson process $i \in \{1, 2\}$ in response to a tone of frequency f octaves is computed according to a gaussian tuning curve with center μ_i,

$$r_i(f) = r_{\max} \exp\left[\frac{(f - \mu_i)^2}{-2\sigma^2}\right], \tag{4.1}$$

where both tuning curves have the same width, σ, and maximum firing rate, $r_{\max}$. The bandwidth, in octaves, of each receptive field, measured at half the peak firing rate, is computed using $B = 2\sigma\sqrt{2 \ln 2}$. In what immediately

follows, we assume without loss of generality that a tone is presented only at the center of one of the two receptive fields, that is, $f \in \{\mu_1, \mu_2\}$, the frequency separation between the tones is $\triangle f = |\mu_2 - \mu_1|$ octaves, and

$$r_i(f) = \begin{cases} r_{\max} & f = \mu_i \\ r_{\max} \exp\left(\frac{\triangle f^2}{-2\sigma^2}\right) & f \neq \mu_i. \end{cases} \tag{4.2}$$

4.1 First Spike Criterion with Ideal Memory. The first spike criterion sets $N = 1$ so that the abstract SSA model assumes that the first counter to register a single spike is the one at which the signal is located. This algorithm, though simple, provides a useful vehicle by which to introduce the expected SI principle.

Let the tones have frequencies $f_A = \mu_1$ and $f_B = \mu_2$. According to the first-spike criterion, the probability that the tone A is registered when tone A is presented is the same as the probability that Poisson process 1 emits a spike first,

$$\begin{aligned} P(\hat{X} = A \mid X = A) &= \int_0^\infty r_{\max} \exp(-r_{\max} t) \exp(-r_{\max} q t)\, dt \\ &= \frac{1}{1+q}, \end{aligned} \tag{4.3}$$

where $q \equiv \exp(\frac{\triangle f^2}{-2\sigma^2})$. Equation 4.3 is the probability that a spike arrives at time t at input A, and at a time later than t at input B, integrated over all positive t. Then, due to symmetry, we also have

$$P(\hat{X} = B \mid X = B) \;=\; P(\hat{X} = A \mid X = A) \;=\; P(\text{correct}).$$

Let us assume that the abstract SSA model has an ideal memory (it knows a priori which tone is the deviant),[2] and it emits K spikes when it sees a deviant and zero spikes otherwise:

$$k'(\hat{x}, \hat{y}) = \begin{cases} K & \hat{x} = \hat{y} \\ 0 & \hat{x} \neq \hat{y}. \end{cases}$$

[2]In terms of the reduced ASSAM formulation given in section 3.2, this means specifying $h(m) = y$ for a known deviant y. As h no longer depends on m, the choice of l' no longer affects the behavior of the model and can be ignored. See also Figure 2.

In that case, the expression for ESI in equation 2.3 reduces (if A is deviant)[3] to

$$\mathrm{ESI} = \frac{K\,P(\hat{X} = A \mid X = A) - K\,P(\hat{X} = A \mid X = B)}{K\,P(\hat{X} = A \mid X = A) + K\,P(\hat{X} = A \mid X = B)} \tag{4.4}$$

$$= \frac{P(\hat{X} = A \mid X = A) - [1 - P(\hat{X} = B \mid X = B)]}{P(\hat{X} = A \mid X = A) + [1 - P(\hat{X} = B \mid X = B)]} \tag{4.5}$$

$$= 2P(\text{correct}) - 1 \tag{4.6}$$

$$= \frac{1-q}{1+q}, \tag{4.7}$$

and the same obtains if one starts with B as the deviant. Note that the ESI does not depend on the overall input rate, $r_{\max}$, or the output firing rate K, but it does depend on the ratio of $\triangle f$ and σ (or B) via q. Specifically, as the tones move further apart or the peripheral bandwidth decreases, the ESI rises. This is consistent with the experimental finding that increasing $\triangle f$ increases SI (see section 2). We defer plots showing how ESI varies with $\triangle f/\sigma$ until the next section, where the above result is obtained in a more general context.

4.2 *N*-Spike Criterion with Ideal Memory. The N-spike criterion waits for one of the counters to receive N spikes before a decision is made, thereby generalizing the first-spike criterion described above. The probability that the time, T_N, of the Nth spike from Poisson process i falls in the vanishing interval $(t, t + \delta t]$ is given by the probability density function of the Erlang distribution (Papoulis, & Pillai, 2002), namely,

$$P(t < T_N \leq t + \delta t) = \frac{r_i^N t^{N-1}}{(N-1)!} \exp(-r_i t)\ \delta t, \quad t \geq 0, \delta t \to 0. \tag{4.8}$$

The gamma distribution is a generalization of the Erlang distribution, in which N can assume continous values. The suitability of the gamma distribution for describing the time it takes for a neuron to reach threshold given Poisson inputs has been previously discussed by Stein (1965). A model of this kind has also recently been used by Oster, Douglas, and Liu (2009) in a mathematical treatment of winner-take-all networks.

[3]For brevity, we will often provide derivations only for the case where A is presented as the deviant, when the derivation for the other three eventualities (A as standard; B as deviant and standard) proceeds along exactly the same lines and requires no additional insight.

The probability that it takes Poisson process i strictly longer than t seconds to emit N spikes is given by the complement of the Erlang cumulative distribution function (c.d.f.), that is,

$$P(T_N > t) = \sum_{n=0}^{N-1} \frac{(r_i t)^n}{n!} \exp(-r_i t), \quad t \geq 0. \tag{4.9}$$

If we again make the association $f_A = \mu_1$ and $f_B = \mu_2$, then, according to the N-spike criterion, the probability that tone A is registered when tone A is presented is the same as the probability that Poisson process 1 emits N spikes before process 2 does. This is equivalent to the probability that process 1 emits its Nth spike at time t and process 2 takes longer than t to emit N spikes, integrated over all t. Accordingly, forming the product of equations 4.8 and 4.9 and performing the integration,

$$\begin{aligned}
&P(\hat{X} = A \mid X = A) \\
&= \int_0^\infty \frac{r_{\max}^N t^{N-1}}{(N-1)!} \exp(-r_{\max} t) \sum_{n=0}^{N-1} \frac{(r_{\max} q\, t)^n}{n!} \exp(-r_{\max} q\, t)\, dt \\
&= \frac{r_{\max}^N}{(N-1)!} \sum_{n=0}^{N-1} \frac{(r_{\max} q)^n}{n!} \int_0^\infty t^{n+N-1} \exp(-r_{\max}(1+q)t)\, dt \\
&= \sum_{n=0}^{N-1} \frac{(n+N-1)!}{n!(N-1)!} \cdot \frac{q^n}{(1+q)^{n+N}} \\
&= \sum_{n=0}^{N-1} \binom{n+N-1}{n} \cdot \frac{q^n}{(1+q)^{n+N}},
\end{aligned} \tag{4.10}$$

where $\binom{\cdot}{\cdot}$ denotes a binomial coefficient.

Due to symmetry, $P(\hat{X} = B \mid X = B) = P(\hat{X} = A \mid X = A) = P(\text{correct})$. By placing equation 4.10 in 2.3 and following the same procedure as in equations 4.4 to 4.6 a reduced ASSAM with ideal memory that emits K output spikes in response to a deviant and zero spikes in response to a standard is shown to have

$$\text{ESI} = 2 \sum_{n=0}^{N-1} \binom{n+N-1}{n} \cdot \frac{q^n}{(1+q)^{n+N}} - 1, \tag{4.11}$$

where, again, the ESI depends on neither the overall input rate $r_{\max}$ nor output rate K. Note that all these results reduce to those for the first spike criterion when $N = 1$.

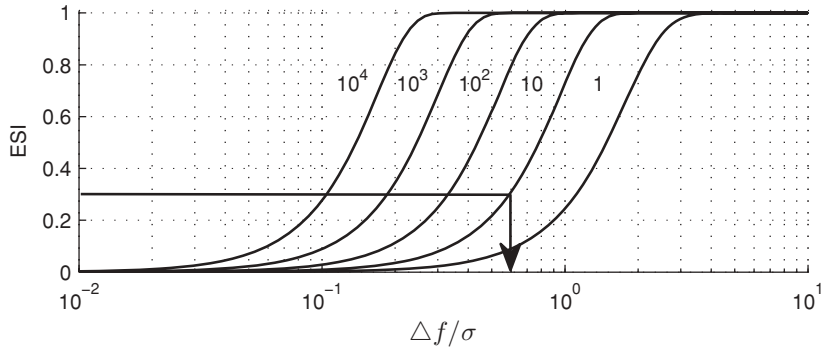

Figure 3: Expected SI for the N-spike criterion with ideal memory as a function of $\triangle f/\sigma$ for various N. For $N \in \{1, 10\}$, the curves were computed using equation 4.11; for $N \in \{100, 1000, 10{,}000\}$, the curves were computed using equation 4.12. The arrow reads off the $N = 10$ curve the ratio $\triangle f/\sigma$ required to obtain an ESI of 0.31—the value obtained by Malmierca et al. (2009) for $\triangle f = 0.15$ octaves (see text).

For large N, equation 4.11 becomes harder to evaluate in practice due to the need to compute an explicit summation and a series of binomial coefficients. A solution for large N is to approximate the Erlang distribution for rate r_i with a gaussian with mean N/r_i and standard deviation $\sqrt{N}/r_i$. Then, retracing the same procedure as above, we obtain

$$\mathrm{ESI} \approx 2\Phi\left(\sqrt{N} \cdot \frac{1-q}{\sqrt{1+q^2}}\right) - 1, \tag{4.12}$$

where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution.

Figure 3 shows a family of curves plotting the ESI for the N-spike criterion with ideal memory as a function of $\triangle f/\sigma$. Each curve demonstrates clearly a monotonically increasing relationship between the separation of the tones ($\triangle f$) and ESI. Similarly, narrowing the peripheral bandwidth by decreasing σ raises the ESI. Finally, using a larger value for N increases the ESI because a longer time period is permitted to gather evidence before deciding how many spikes to respond with.

These considerations imply that if the circuit must react within a certain time period, the maximum spike rate ($r_{\max}$) is fixed, and the tone separation and peripheral bandwidths (and hence their ratio) are fixed, then an upper bound on the ESI is established, which has implications for computational models.

As an example, consider how a reduced, binary ASSAM using the N-spike criterion might reproduce a result in Malmierca et al. (2009). A

two-tone oddball stimulus was presented with $\triangle f = 0.15$ octaves and $p_{dev} = 0.1$.[4] The SI value averaged over 122 neurons in the guinea pig inferior colliculus was 0.31, and neurons tended to respond transiently within about 25 ms of the stimulus onset. If we assume that $r_{\max} = 400$ spikes per second in the model, then in 25 ms, the Poisson process on which the tone is centered will generate, on average, 10 spikes. From Figure 3, for $N = 10$, to secure an ESI of 0.31, a ratio of approximately $\triangle f/\sigma = 0.6026$ is required. With $\triangle f = 0.15$, the largest permissible value for σ is 0.2489, which corresponds to a peripheral bandwidth of approximately 0.59 octaves. This is on the same order of magnitude as the tuning of auditory peripheral neurons in mammals (e.g., Liberman, 1978, in cat).

4.3 Standard-as-Mode Memory Criterion. Thus far in this section, we have assumed that the ASSAM has an ideal memory, that is, it is informed (correctly) of which tone is the deviant. In practice, a full-blown account of stimulus-specific adaptation—abstract or concrete—will involve learning what a deviant is by observing input patterns. Furthermore, if, after a few presentations, the deviant and standard tones swap roles, then SSA is expected to undergo a reversal, as demonstrated in the switching-blocks paradigm (Ulanovsky et al., 2004).

In this section, we shall examine how the ESI is maximized when the ASSAM has access to a memory holding the $L \geq 0$ most recently classified tones. If the arriving tone is t_n, then the memory m is the list $\mathsf{t}_{n-1}, \mathsf{t}_{n-2}, \ldots, \mathsf{t}_{n-L}$. The standard-as-mode criterion is stated as follows: using $|x|_m$ to denote the number of tones of type $x \in \{A, B\}$ in the memory m,

$$h(m) = \begin{cases} A & |B|_m > |A|_m \\ B & |A|_m > |B|_m, \\ \text{uniformly random } A \text{ or } B & |A|_m = |B|_m \end{cases} \tag{4.13}$$

where $h(m)$ is the function that estimates what the deviant is from the memory. The memory update function $l'(\hat{x}, m)$ removes the tail of the list m and concatenates $\hat{x}$ to its head.

We now consider the ESI if the standard-as-mode criterion is applied in conjunction with some estimation criterion (e.g., N-spike). Using ϵ to denote the incoming spike data, let the ASSAM emit K spikes if and only if $g(\epsilon) = h(m)$ and zero otherwise; that is, it emits spikes only if its estimate

[4] In Malmierca et al. (2009), following Ulanovsky et al. (2004, 2003), tone separation was computed according to the normalized frequency difference metric $\triangle f_{norm} \triangleq \frac{f_2 - f_1}{\sqrt{f_1 f_2}}$. The tone separation given in the letter was actually $\triangle f_{norm} = 0.1$, which is roughly equivalent to $\triangle f = 0.15$ octaves. The procedure for converting between the normalized frequency difference and octave frequency difference is provided in the appendix.

of the incoming tone and the deviant tone match. The expected number of spikes in response to a presentation of $x \in \{A, B\}$ as deviant or standard is then derived from equation 3.2 or 3.3, respectively, that is,

$$E\{D_x\} = \sum_{\hat{x}\in\{A,B\}} K\,P(\hat{X} = \hat{x} \mid X = x)P(\hat{y} = \hat{x} \mid Y = x) \tag{4.14}$$

$$E\{S_x\} = \sum_{\hat{x}\in\{A,B\}} K\,P(\hat{X} = \hat{x} \mid X = x)P(\hat{y} = \hat{x} \mid Y \neq x). \tag{4.15}$$

Note that the ASSAM sometimes responds correctly to a tone because its estimates of both the incoming tone and the current deviant are in error.

The expressions of the form $P(\hat{y} = \hat{y} \mid Y = y)$, with $y, \hat{y} \in \{A, B\}$, represent the probability that the function $h(m)$ returns $\hat{y}$ when y is actually deviant. To derive this quantity, we consider how tones are added to the memory. For instance, the probability that l adds A to the memory given A is the deviant is

$$\begin{aligned} P(\hat{X} = A \mid Y = A) &= P(\hat{X} = A \mid X = A)P(X = A \mid Y = A) \\ &\quad + P(\hat{X} = A \mid X = B)P(X = B \mid Y = A) \\ &= P(\hat{X} = A \mid X = A)p_{dev} \\ &\quad + P(\hat{X} = A \mid X = B)(1 - p_{dev}). \end{aligned} \tag{4.16}$$

Because the tones that arrive are statistically independent and the classification of tones is independent of m, $|A|_m$ is governed by a binomial distribution (Papoulis et al., 2002),

$$\begin{aligned} P(|A|_m = n) &= \binom{L}{n} P(\hat{X} = A \mid Y = A)^n [1 - P(\hat{X} = A \mid Y = A)]^{L-n} \\ &= \binom{L}{n} P(\hat{X} = A \mid Y = A)^n P(\hat{X} = B \mid Y = A)^{L-n}, \\ P(|B|_m = n) &= P(|A|_m = L - n), \end{aligned} \tag{4.17}$$

where $\binom{L}{n}$ denotes the binomial coefficient, L-choose-n. The probability that $h(m)$ returns A is the probability that fewer than half of the tones in m are A,

$$P(\hat{y} = A \mid Y = A) = \begin{cases} \sum_{n=0}^{\lceil L/2\rceil - 1} P(|A|_m = n) & L \text{ odd} \\ \sum_{n=0}^{\lceil L/2\rceil - 1} P(|A|_m = n) & \\ \qquad + \frac{1}{2} P(|A|_m = \frac{L}{2}) & L \text{ even}, \end{cases} \tag{4.18}$$

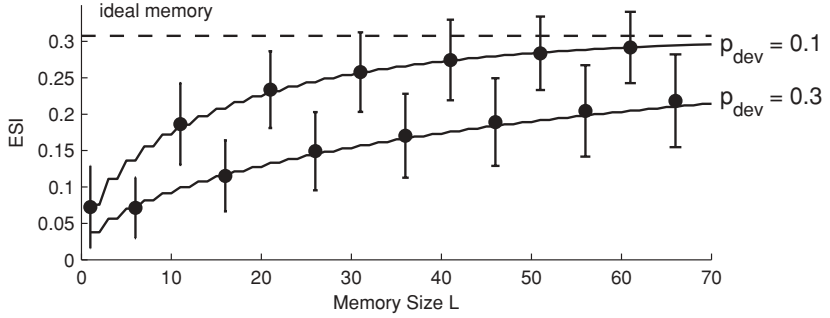

Figure 4: Combining the N-spike ($N = 10$) and standard-as-mode criteria. Solid curves plot the analytical ESI against memory size L for two deviant probabilities, 0.1 and 0.3. The dashed line plots the ESI for an ideal memory. Solid markers show the empirical ESI averaged over 1000 simulations of an 800-tone oddball sequence. Error bars denote the standard deviation of the ESI. In this example, all the curves are specific to $\triangle f/\sigma = 0.6026$—a value motivated by results in Malmierca et al. (2009) (*cf.* section 4.2).

where $\lceil \cdot \rceil$ rounds up to the nearest integer, and similar results obtain for the other three possibilities. By replacing equation 4.16 in 4.17 in 4.18 and so on, and then substituting these results into equations 4.14 and 4.15 and finally into equation 2.3, one arrives at the expected SI.

Before moving on, note that when the system is memoryless, $L = 0$, then $\forall y \in \{A, B\}$, $P(\hat{y} = y \mid Y = y) = \frac{1}{2}$, which implies that ESI $= 0$. On the other hand, provided that tone classification outperforms a chance decision, as $L \to \infty$, $P(\hat{y} = y \mid Y = y) \to 1$, and the ESI tends toward its value for ideal memory.

The example given in section 4.2, motivated by a particular experimental result in Malmierca et al. (2009), can be augmented to include the standard-as-mode criterion for memory, as opposed to ideal memory. Figure 4 plots the ESI against memory length, L, for two deviant probabilities, $p_{dev} = 0.1, 0.3$; $\triangle f = 0.15$ octaves, $\sigma = 0.25$ (equivalent bandwidth = 0.59 octaves); $N = 10$. The solid markers plot the mean (empirical) ESI obtained from 1000 simulations, each with an oddball sequence of 800 tones, identical to the overall length of the oddball sequences used by Malmierca et al. (2009). The error bars indicate the standard deviation around these means.

The following observations may be made in connection with Figure 4. First, the staircase-like appearance of the curves plotting the analytical ESI is not artifactual, but rather stems from the fact that the ESI for an even memory of size $L > 0$ equals that for a memory of size $L - 1$. Second, increasing the memory size causes the ESI to converge toward that obtained for an ideal memory. Naturally, giving the ASSAM access to a longer tone history increases the accuracy of its decision concerning the deviant tone.

Third, increasing the rarity of deviants reliably increases the ESI for a given L, as illustrated by the two curves in Figure 4. (The principle holds for arbitrary p_{dev}; results not shown.) The wider implications of these findings are discussed in section 7.

4.4 Depressing-Standard Memory Criterion. The standard-as-mode criterion explicitly infers the deviant from a buffer of recent tones, assigning equal weight to each element. A more natural way to conceive of a biological system storing recent stimuli is in the form of a dynamical system whose state is modified by each tone as it is presented. For example, Tsodyks and Markram (1997) have presented a computational model of a synapse that depresses following the arrival of each presynaptic spike. In this section, we assume that the presentation of a tone leads to the depletion of some stimulus-specific synaptic resources, so that the repetition of a standard tone leads to reduced spiking. Consequently, we designate this approach the *depressing-standard* criterion.

Let the memory m of an ASSAM consist of a pair of state variables (m_A, m_B), initially set to (1, 1). If an incoming tone is identified as $\hat{x} \in \{A, B\}$ on the basis of input ϵ, then let the number of output spikes be proportional to $m_{\hat{x}}$,

$$k(\hat{x}(\epsilon), m) = K\, m_{\hat{x}},\ \ K > 0. \tag{4.19}$$

After k has been computed, let each m_x, $x \in \{A, B\}$, in the memory m be updated according to the rule

$$m_x = \begin{cases} \alpha m_x & x = \hat{x} \quad \text{(adaptation)} \\ m_x + \beta(1 - m_x) & x \neq \hat{x} \quad \text{(recovery)}, \end{cases} \tag{4.20}$$

where $\alpha, \beta \in (0, 1)$. (This constitutes the function l.) Thus, if tone A arrives, m_A adapts and m_B recovers and vice versa. The rates of adaptation and recovery are controlled by the parameters α and β, respectively.

The ESI in equation 2.3 is solely a function of terms $E\{D_A\}$, $E\{S_A\}$, $E\{D_B\}$, and $E\{S_B\}$. These terms are in turn related to the expected values of m_x. For example, the expected number of spikes for tone A when presented as deviant is

$$\begin{aligned} E\{D_A\} &= E\left\{ \sum_{\hat{x} \in \{A,B\}} k(\hat{x}, m) P(\hat{X} = \hat{x} \mid X = A) \mid Y = A \right\} \\ &= \sum_{\hat{x} \in \{A,B\}} K\, E\{m_{\hat{x}} \mid Y = A\} P(\hat{X} = \hat{x} \mid X = A), \end{aligned} \tag{4.21}$$

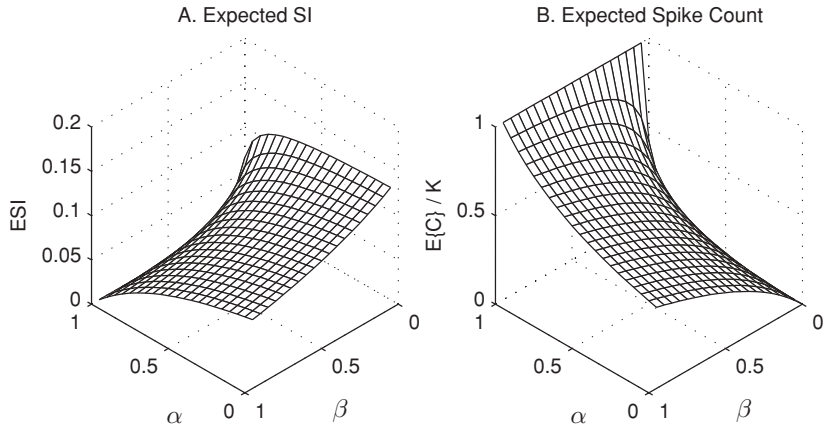

Figure 5: Surfaces plotting the (A) expected SI and (B) expected spike count ($E\{C\}$) for various choices of adaptation (α) and recovery (β) parameters, with $\triangle f/\sigma = 0.60$, $p_{\text{dev}} = 0.1$ and $N = 10$. The ESI for an ideal memory is 0.3075.

with similar results following for $E\{S_A\}$, $E\{D_B\}$ and $E\{S_B\}$, mutatis mutandis. To evaluate the ESI with A as the deviant thus requires one to evaluate $E\{m_A \mid Y = A\}$ and $E\{m_B \mid Y = A\}$. Concentrating on the former, we see that

$$E\{m_A \mid Y = A\} = E\{\alpha m_A \mid Y = A\} P(\hat{X} = A \mid Y = A) + E\{m_A + \beta(1 - m_A) \mid Y = A\} P(\hat{X} = B \mid Y = A). \tag{4.22}$$

Dividing both sides of equation 4.23 by $E\{m_A \mid Y = A\}$ and rearranging yields

$$E\{m_A \mid Y = A\} = \frac{\Big[1 - P(\hat{X} = A \mid Y = A)\Big]\beta}{P(\hat{X} = A \mid Y = A)(1 - \alpha - \beta) + \beta}. \tag{4.23}$$

The evaluation of $P(\hat{X} = A \mid Y = A)$ has already been addressed in equation 4.16. With the expectations $E\{m_B \mid Y = A\}$, $E\{m_A \mid Y = B\}$, and $E\{m_B \mid Y = B\}$ evaluated using the same principle as for $E\{m_A \mid Y = A\}$, demonstrated above, the quantities $E\{D_A\}$, $E\{S_A\}$, $E\{D_B\}$, and $E\{S_B\}$, and hence also the ESI, follow directly.

Figure 5A is a surface plot showing the effect of varying α and β on the ESI. From this plot, it is clear that the ESI monotonically increases as both α and β decrease, such that the maximum ESI is approached as $\beta \to 0$ with $\alpha = 0$. Under such circumstances, the adaptation following each tone

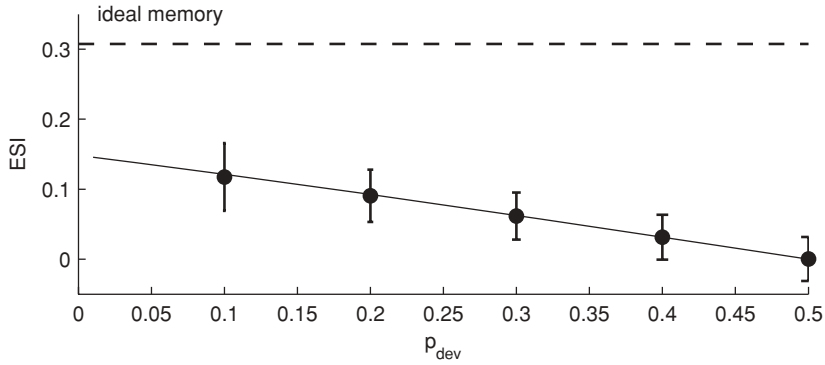

Figure 6: A graph plotting the ESI against the probability of a deviant (solid), overlaid with markers and error bars that plot, respectively, the mean and standard deviation of the ESI obtained in simulations (see text for details). The dashed line plots the ESI for an ideal memory.

is total, and the recovery is infinitely slow (though non zero). Conversely, setting $\alpha = 1$ implies a lack of adaptation, in which case there can be no SSA.

Two further remarks are appropriate at this juncture. First, whereas letting $L \to \infty$ in the standard-as-mode criterion causes the ESI to approach the value for an ideal memory, the same does not hold for the depressing-standard criterion as $\alpha, \beta \to 0$; in fact, the maximum ESI (≈ 0.14) is less than half that obtained for an ideal memory (≈ 0.31). Second, although a reasonably large ESI can still be achieved by a rapid adaptation and slow recovery, one should note that it comes at the expense of a decline in the overall responsiveness of the neuron. Figure 5B plots the normalized expected spikes per tone as a function of α and β and bears out the intuition that very slow recovery rates are associated with very little spiking activity. Another interpretation of this result is that by setting β to a value near zero, one configures the unit to generate spikes only for deviants that are preceded by a vast number of standards. This is an example of what we call hedging: a situation in which an ASSAM secures a high ESI by emitting spikes only when it concludes that the tone is almost certainly a deviant. We shall examine this behavior more closely in section 6.

Figure 6 fixes $\alpha = 0.5$ and $\beta = 0.1$ and plots the ESI as a function of the probability of a deviant (solid line). As with the standard-as-mode criterion, the ESI decreases as deviants become more common, with ESI $= 0$ for $p_{dev} = 0.5$. Markers show the mean and standard deviation of the SI obtained over the course of 10,000 simulations. Each simulation measured the response of the model to an oddball sequence of 800 tones, in which $\triangle f = 0.15$ octaves, $\sigma = 0.25$ (equivalent bandwidth = 0.59 octaves), and $N = 10$.

5 SSA with Multiple Poisson Inputs

So far we have confined our attention to networks with two Poisson input neurons, the tuning curves of which have been centered on the two frequencies that appear in the oddball sequence; consequently, the model is to some extent hard-wired for particular stimuli. However, in a biological neuron, SSA is observed for various pairs of tone frequencies within the receptive field. Physiological experiments generally use a pair of tones equally spaced on either side of the neuron's best frequency on an octave scale. This section shows that the analytical techniques for predicting the ESI for two-input ASSAMs described above generalize to networks with M inputs.

The fact that more than two tones can now be presented to the network requires some mild modification to the nomenclature. Rather than referring to A and B as the random events corresponding to stimulation by tone A or B, there are now M tones that can be presented, labeled X_i, $i \in \{1, \ldots, M\}$. When presented with a tone, the network always classifies it as exactly one of these, so we also have $\hat{X}_i$, $i \in \{1, \ldots, M\}$, where $\sum_{i=1}^{M} P(\hat{X}_i) = 1$. Finally, we introduce the random variable Z to denote the standard tone, just as Y has denoted the deviant. (In the two-input networks considered earlier, the standard did not need to be specified explicitly, as it was implied by the deviant.)

5.1 Extending the *N*-spike and Depressing-Standard Criteria. In this section, we describe how the N-spike and depressing-standard criteria can be extended to handle an arbitary number of inputs. The reasoning behind each step does not differ from the two-input case, which affords us some brevity in the derivation that follows, but the combinatorial explosion of possibilities (e.g., in spike arrival times) does sometimes lead to somewhat unwieldy expressions for the probabilities involved.

Let $\hat{X} = X_j$, $j \in \{1, \ldots, M\}$, denote the random event that Poisson process j is the first of all processes to emit N spikes and, on this basis, that the ASSAM concludes that the tone is positioned at frequency f_{X_j}. The probability that an ASSAM identifies $\hat{X}_j$ when the tone is actually located at X_k is

$$P(\hat{X} = X_j \mid X = X_k) = \int_0^\infty \frac{r_j^N}{(N-1)!} t^{N-1} \exp(-r_j t) \prod_{i \neq j}^{M} \sum_{n=0}^{N-1} \frac{(r_i t)^n}{n!} \exp(-r_i t)\, dt, \tag{5.1}$$

where $r_i \equiv r_i(k)$. Note that equation 5.1 is the probability that the Nth spike at unit j arrives at time t, together with the probability that the Nth spikes at all other units $i \neq j$ arrive later than t, integrated over all possible t.

The integral in 5.1 can be solved (see appendix) to give the closed-form expression

$$P(\hat{X} = X_j \mid X = X_k) = \sum_{n_1=0}^{N-1} \cdots \sum_{n_{j-1}=0}^{N-1} \sum_{n_{j+1}=0}^{N-1} \cdots \sum_{n_M=0}^{N-1} \frac{\left(N-1+\sum_{i\neq j}^{M} n_i\right)! \, r_j^N \prod_{i\neq j}^{M} r_i^{n_i}}{(N-1)! \prod_{i\neq j}^{M} n_i! \left(\sum_i^M r_i\right)^{(N+\sum_{i\neq j}^{M} n_i)}} \quad (5.2)$$

(where again the dependence of r_i on k has been suppressed for brevity). Evaluating equation 5.2 for all j and k generates a confusion matrix for all possible tone presentations and classifications, with correct decision probabilities along the diagonal.

The depressing-standard criterion is based on the idea that a frequently presented tone will deplete some stimulus-specific resources and cause a reduction in spiking. The two-input model described in section 4.4 requires a two-element memory; for an M-input model, we generalize to an M-tuple memory, $m = (m_1, \ldots, m_M)$, which is updated according to the same rule, equation 4.20.

Let $A, B \in \{1, \ldots, M\}$ denote the indices of the two tones that are presented to the ASSAM in an oddball experiment. The four components that constitute the expression for the ESI then assume the values

$$E\{D_{X_A}\} = \sum_{i=1}^{M} E\{m_i \mid Y = X_A, Z = X_B\} P(\hat{X} = X_i \mid X = X_A)$$

$$E\{S_{X_A}\} = \sum_{i=1}^{M} E\{m_i \mid Y = X_B, Z = X_A\} P(\hat{X} = X_i \mid X = X_A)$$

$$E\{D_{X_B}\} = \sum_{i=1}^{M} E\{m_i \mid Y = X_B, Z = X_A\} P(\hat{X} = X_i \mid X = X_B)$$

$$E\{S_{X_B}\} = \sum_{i=1}^{M} E\{m_i \mid Y = X_A, Z = X_B\} P(\hat{X} = X_i \mid X = X_B).$$

The conditional probabilities are evaluated using equation 5.2. The procedure to obtain the expected value of the memory element m_i is identical in principle to that used in the two-input case (see section 4.4). For example, with X_A as the deviant, we obtain the expression

$$E\{m_i \mid Y = X_A, Z = X_B\} = \frac{\left[1 - P(\hat{X} = X_i \mid Y = X_A, Z = X_B)\right]\beta}{P(\hat{X}_i \mid Y = X_A, Z = X_B)(1-\alpha-\beta)+\beta}, \quad (5.3)$$

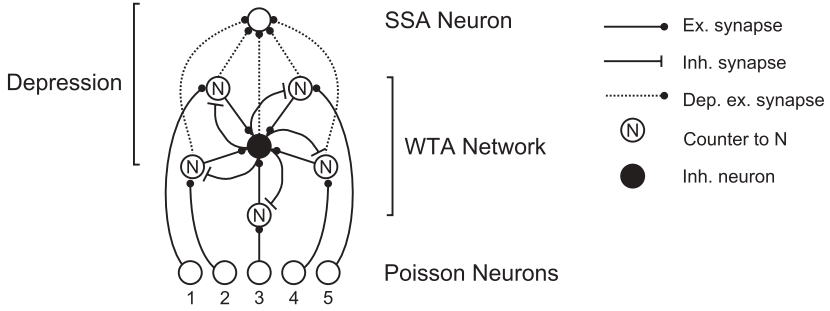

Figure 7: Diagram of a five-input abstract SSA model that implements the N-spike and depressing-standard criteria. The winner-take-all (WTA) section of the network implements the N-spike criterion and is responsible for estimating the incoming tone ($\hat{X}$). The section labeled "depression" implements the depressing-standard criterion.

where

$$P(\hat{X} = X_i \mid Y = X_A, Z = X_B) = P(\hat{X} = X_i \mid X = X_A)p_{dev} + P(\hat{X} = X_i \mid X = X_B)(1 - p_{dev}). \quad (5.4)$$

The probabilities on the right-hand side of equation 5.4 are found using equation 5.2. Setting $M = 2$ reduces all these expressions to the two-input case.

5.2 Simulating a Small Winner-Take-All Network. Having derived, in stages, an expression for the ESI for an M-input network using the N-spike and depressing-standard criteria, we are in a position to simulate the output of a small network and compare the theoretical and empirical values. The abstract SSA model corresponds to the network shown in Figure 7.

In this simulation, the Poisson neurons, labeled 1 to 5 along the bottom, are spaced 0.075 octaves apart and have tuning curves with $\sigma = 0.25$. The units labeled N are counters that await N spikes before firing a single spike; here, we set $N = 10$. Once a single spike has been emitted, the inhibitory (black) unit prevents any unit from firing until the next tone is presented. The dotted arcs represent depressing synapses which adapt and recover according to equation 4.2. Finally, the top neuron emits a number of spikes directly proportional to its input, which, in this scheme, is always via exactly one synapse.

Whereas in the binary network there are only two centers at which to present tones, in this network there are five. Following physiological studies, we presented tones A and B in an oddball sequence on either side of a

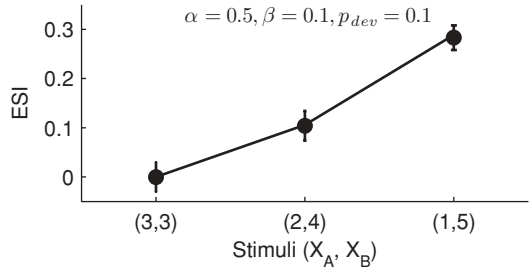

Figure 8: A graph connecting three theoretical ESI values for three tone spacings on either side of the center unit (i.e., 3), overlaid with markers and error bars that plot respectively the mean and standard deviation of the ESI obtained in simulations (see text for details).

Table 1: Table of ESI Values for Tones A and B.

	A = 1	A = 2	A = 3	A = 4	A = 5
B = 1	0	0.0227	0.0941	0.1949	0.2890
B = 2	0.0227	0	0.0288	0.1059	0.1949
B = 3	0.0941	0.0288	0	0.0288	0.0941
B = 4	0.1949	0.1059	0.0288	0	0.0227
B = 5	0.2890	0.1949	0.0941	0.0227	0

nominal "best frequency," corresponding to the middle input: ($A=3$, $B=3$), ($A=2$, $B=4$) and ($A=1$, $B=5$). The analytical and simulation results are plotted in Figure 8 and agree closely—as indeed they should, given that in this case, we were able to derive directly the analytical expressions without the need for approximations. Figure 8 shows that increasing the separation between the tones causes the ESI to rise. More generally, Table 1 lists a matrix including the ESI for all pairs of tones. As one would expect, the matrix is symmetric with zeroes along its main diagonal. The off-diagonals correspond to a particular $\triangle f$. (Pairs of tones presented at the center of the Poisson array are slightly easier to discriminate than those presented at the periphery; consequently, the ESI values are larger toward the center of each diagonal in the matrix.)

5.3 Depressing-Standard Criterion and Entropy. The idea of a sequence of depressing channels implicitly keeping track of the recent stimulus history raises the question of how the overall activity of an SSA neuron varies according to the relative proportions of tones that constitute

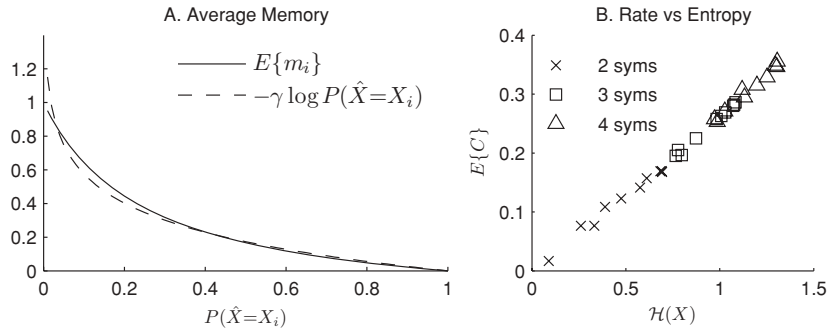

Figure 9: (A) A graph comparing the expected memory value, $E\{m_i\}$ and the (natural) log probability, $-\gamma \log P(\hat{X} = X_i)$, when $\gamma = 0.25$. (B) A scatter plot comparing the oddball sequence entropy to the firing rate of an ASSAM with ideal discrimination and depressing-standard memory. The correlation coefficient taken across all the data points is 0.990 (3 d.p.).

the oddball sequence. For the depressing-standard criterion, the expected spike count per tone for an SSA neuron, $E\{C\}$, is

$$E\{C\} = \sum_{i=1}^{M} P(\hat{X} = X_i)\, E\{m_i\}. \tag{5.5}$$

The expression for the expected value of the memory m_i given an oddball sequence that leads to tone estimation probabilities $P(\hat{X} = X_i)$, $i \in \{1, \ldots, M\}$, is

$$E\{m_i\} = \frac{(1 - P(\hat{X} = X_i))\beta}{P(\hat{X} = X_i)(1 - \alpha - \beta) + \beta}. \tag{5.6}$$

For the choice of parameter values $\alpha = 0.5$ and $\beta = 0.1$, which we have been using so far, the shape of $E\{m_i\}$ as a function of $P(\hat{X} = X_i)$ appears approximately directly proportional to the negated logarithm of $P(\hat{X} = X_i)$, as Figure 9A shows. That is, $E\{m_i\} \approx -\gamma \log P(\hat{X} = X_i)$, for some positive constant of proportionality γ, and we can write

$$E\{C\} \approx -\gamma \sum_{i=1}^{M} P(\hat{X} = X_i) \log P(\hat{X} = X_i) = \gamma\, \mathcal{H}(\hat{X}) \tag{5.7}$$

where $\mathcal{H}(\hat{X})$ denotes the entropy of the estimations. Informally, then, the average firing rate of the model is approximately proportional to its uncertainty regarding which tone will be estimated next. To relate the input

to the output more directly, the entropy of $\hat{X}$ can be expanded into two terms:

$$\mathcal{H}(\hat{X}) = \mathcal{H}(X) + \sum_{i,j=1}^{M} P(\hat{X} = X_i, X = X_j) \log \left[\frac{P(X = X_j)}{P(\hat{X} = X_i)} \right].$$

The first term, $\mathcal{H}(X)$, denotes the entropy of X and reflects the uncertainty concerning the oddball sequence itself. The second term corresponds to any additional uncertainty introduced by the estimation procedure. This latter term is zero when discrimination is ideal, so that the firing rate is approximately proportional to the entropy of the oddball sequence: $E\{C\} \approx \gamma \mathcal{H}\{X\}$.

This final point can be demonstrated empirically by synthesizing oddball sequences with various entropies and calculating the average firing rate of the ASSAM with ideal discrimination and a depressing-standard memory. Figure 9B shows some example data for three sets of 10 oddball sequences: the first set contains two symbols (A and B); the second set contains three symbols; the third set contains four symbols. The large correlation coefficient between the entropy and firing rate ($\rho > 0.990$) supports the main result of this section: that the output activity is approximately proportional to stimulus entropy. Naturally, the choice of α and β has an impact on the validity of the logarithmic approximation and, hence, the conclusions regarding entropy that follow. But log-proportional curves appear to approximate $E\{m_i\}$ quite well, provided that $\alpha \leq 0.5$ and $\beta \approx 0.1$ (that is, if adaptation is rapid and recovery is slow), especially in the regions of least curvature, where $P(\hat{X} = X_i) > 0.5$ (results not shown).

5.4 Receptive Fields for the First-Spike Criterion. Thus far we have extended the two-input model to an M-input model. In this section, we proceed a step further and examine the possibility of deriving a field model, which replaces the discrete units (e.g., the Poisson neurons and counters indexed i previously) with functions of continuous variables.

The first-spike criterion passes the first, and only the first, spike to emerge from an array of M units. Taking equation 5.2 and setting $N = 1$, the probability that the symbol X_j is estimated when the stimulus is X_k is presented is

$$P(\hat{X} = X_j \mid X = X_k) = \frac{r_j(k)}{\sum_i^M r_i(k)}. \tag{5.8}$$

Instead of computing the probability that unit j emits the first spike when a tone is presented at unit k, we associate a probability density with the first spike arriving at location $\hat{f}$, given that the tone is presented at frequency

f. Thus, f and $\hat{f}$ serve as continuous analogs of X and $\hat{X}$, and a continuous version of equation 5.8 is obtained by replacing the summation in the denominator with an integral. Then, as the form of $r(\cdot)$ from equation 4.2 is a gaussian, we have

$$p(\hat{f} \mid f) = \frac{r_{\max} \exp\left[\frac{(f-\hat{f})^2}{-2\sigma^2}\right]}{\int_{-\infty}^{\infty} r_{\max} \exp\left[\frac{(f-s)^2}{-2\sigma^2}\right] ds} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(f-\hat{f})^2}{-2\sigma^2}\right]. \tag{5.9}$$

When a tone is identified as having frequency $\hat{f}$, we need to generate output spikes and modify the memory (either cause it to adapt or allow it recover) as in the discrete case. Formerly, a single memory unit m_i was associated with each symbol X_i. In the continuous case, the memory is a field, $m(s)$, where the parameter s serves as the continuous counterpart of the indexing variable i. Now that the probability of an estimation corresponding to any given $\hat{f}$ is vanishingly small, we must assume that the arrival of a spike produces a spread of activity around the point $\hat{f}$. Specifically, let the probability that the first pre synaptic spike arriving at $\hat{f}$ is communicated across the synapse at location s be denoted $P(u(s) \mid \hat{f})$. Then, given that the stimulus is presented at frequency f, we have

$$P(u(s) \mid f) = \int_{-\infty}^{\infty} P(u(s) \mid \hat{f}) p(\hat{f} \mid f)\, d\hat{f}. \tag{5.10}$$

If, for convenience, we choose a gaussian in s with unit peak, mean $\hat{f}$, and standard deviation σ_u for $P(u(s) \mid \hat{f})$, then equation 5.10 admits the explicit solution

$$P(u(s) \mid f) = \frac{\sigma_u}{\sqrt{\sigma^2 + \sigma_u^2}} \exp\left[\frac{(s-f)^2}{-2(\sigma^2 + \sigma_u^2)}\right]. \tag{5.11}$$

This function gives the probability that a spike traverses the synapse at position s, given that a tone is presented at f. However, in the depressing-standard criterion, the efficacy of the presynaptic spike is modulated by the recent history of activity at the synapse. The memory of the synapse at s is denoted $m(s)$, and, in the continuous model, the expected value of the memory is

$$E\{m(s)\} = \frac{[1 - P(u(s))]\,\beta}{P(u(s))(1 - \alpha - \beta) + \beta}, \tag{5.12}$$

where

$$P(u(s)) = P(u(s) \mid f_D)\, p_{dev} + P(u(s) \mid f_S)(1 - p_{dev})$$

and f_D and f_S are the frequencies of the deviant and standard, respectively.

Finally, the expected output of the SSA neuron given input f is an integral over the product of the expected activity at s, $P(u(s) \mid f)E\{m(s)\}$, and a gaussian profile with standard deviation σ_k, namely,

$$E\{k \mid f\} = \int_{-\infty}^{\infty} P(u(s) \mid f)E\{m(s)\} \exp\left(\frac{s^2}{-2\sigma_k^2}\right) ds, \tag{5.13}$$

where the center of the response field of the SSA neuron is set to zero. While the integrand in equation 5.13 has a closed-form solution for any set of parameters, the integral itself does not, on account of the ratio of gaussian mixtures in $E\{m(s)\}$, from equation 5.12. Consequently, we must resort to numerical methods to solve it. (In what follows, we use trapezoidal integration.)

Figure 10 plots the results of the various stages described above (in rows) for various parameter sets (in columns). In column A, the tones are closely spaced ($\triangle f = 0.15$), and the resolution of the Poisson array is quite low (the bandwidth is 0.6 octaves, or $\sigma \approx 0.25$). We assume that the lower-frequency value corresponds to the standard tone and $p_{dev} = 0.1$. The bottom-left plot shows that the overall receptive field adapts everywhere, but develops a slight "shoulder" on the low-frequency side due to the prevalence of standards. This effect is emphasized in column B by moving the tones apart to $\triangle f = 0.5$ octaves and further still in column C by halving the peripheral bandwidth to 0.3 octaves, so that the shoulder in the receptive field develops into a notch. Ulanovsky et al. (2003, Figure 2e) also suggest the development of a receptive field with notches located at standard frequencies. The origin of this particular shape can be traced to the depressed state of the memory at frequencies that are presented often. In column D, increasing the probability of a deviant to $p_{dev} = 0.3$ causes the notch associated with the standard to grow shallower and a second notch to appear, associated with the deviant. In this way, we see that the number of tones that a single unit embedded in a network is able to keep track of depends chiefly on its ability to resolve distinct frequencies.

5.5 SSA and Fisher Information. In section 5.4, we substituted an essentially discrete stimulus parameter (i.e., $f \in \{\mu_1, \mu_2, \ldots\}$) for a continuous one ($f \in \mathbb{R}$). Moreover, we replaced a number of discrete probabilities, connected with the arrival of the first spike at particular units, with a probability

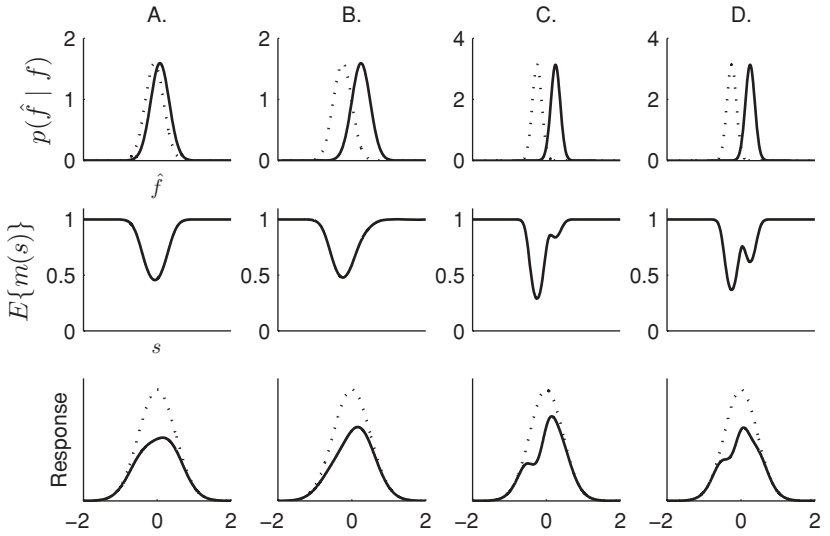

Figure 10: Deriving the expected receptive field of a model SSA neuron. Rows: (1) p.d.f.s governing the spatial location of the first spike to arrive for the deviant (solid) and standard (dotted) tones; (2) average stimulus-specific adaptation of the unit; and (3) expected receptive field of the unit with adaptation (solid) and without (dotted). Columns: (A) bandwidth = 0.6 octaves, $\triangle f = 0.15$ octaves, $p_{dev} = 0.1$; (B) as in *A*, with $\triangle f = 0.5$ octaves; (C) as in *B*, with bandwidth = 0.3 octaves; and (D) as in *C*, with $p_{dev} = 0.3$. In all columns, $\alpha = 0.5$, $\beta = 0.1$, $N = 1$.

density function, governing the distribution of the location of the first spike along a spatial axis.

We start here by noting from equation 5.9 that $\hat{f}$ can in fact be interpreted as an unbiased estimator of f with variance σ^2, that is, $E\{\hat{f}\} = f$ and $E\{(\hat{f} - f)^2\} = \sigma^2$. This prompts the question of whether other unbiased estimators of f—in particular, those with lower variances—can serve in place of the first-spike location statistic. With this in mind, a Fisher information analysis can be employed to establish a lower bound on the variance of an unbiased estimator, $\hat{f}$, based on any kind of evidence, ϵ, via the Cramér-Rao bound.

To take a common example, suppose that the evidence, ϵ, available to the ASSAM consists of the number of spikes received by each element in an array of M Poisson inputs in a time window of T seconds. Let n_i denote the number of spikes received from Poisson process i; the Fisher information delivered by the measurement n_i concerning the stimulus parameter f is

then

$$J_i(f) = E_{n_i}\left\{-\frac{\partial^2}{\partial f^2}\left\{\ln p(n_i \mid f)\right\}\right\}$$
$$= \frac{Tr_{\max}(f-\mu_i)^2}{\sigma^4}\exp\left[\frac{(f-\mu_i)^2}{-2\sigma^2}\right], \tag{5.14}$$

and as the Fisher information is additive for statistically independent random variables, the information conveyed by the entire array is

$$J_{1\ldots M}(f) = \sum_{i=1}^{M}\frac{Tr_{\max}(f-\mu_i)^2}{\sigma^4}\exp\left[\frac{(f-\mu_i)^2}{-2\sigma^2}\right]. \tag{5.15}$$

If we assume an infinite array of equally spaced, broadly overlapping Poisson inputs, such that there are ρ inputs per octave, then one can approximate the overall Fisher information in the evidence ϵ by replacing the summation with an integral:

$$J_\epsilon(f) \approx \rho\int_{-\infty}^{\infty}\frac{Tr_{\max}(f-\mu)^2}{\sigma^4}\exp\left[\frac{(f-\mu)^2}{-2\sigma^2}\right]d\mu$$
$$= \frac{\sqrt{2\pi}\,\rho Tr_{\max}}{\sigma} = J_\epsilon. \tag{5.16}$$

The minimum variance for the unbiased estimator $\hat{f}$ based on ϵ is $1/J_\epsilon$. Thus while the Fisher information approach does not produce the conditional distribution $p(\hat{f} \mid f)$, it does supply us with its first two moments. Intuitively, the distribution is also unimodal and symmetric, so a suitable approximation is the gaussian with p.d.f.:

$$p(\hat{f} \mid f) = \frac{1}{\sqrt{2\pi/J_\epsilon}}\exp\left[\frac{(\hat{f}-f)^2}{-2/J_\epsilon}\right]. \tag{5.17}$$

This probability density function can be substituted into any of the methods we have already described; for instance, equation 5.17 can be used in place of equation 5.9. In the concluding example of section 6, we shall use this type of analysis to estimate the frequency presented to an array of Poisson neurons.

6 Explicit Maximization of the ESI

In earlier sections we discussed reduced abstract SSA models—those that make a binary decision about which tone was presented before responding.

An alternative approach is to maximize the ESI directly on the basis of evidence ϵ. Recall that the function $k(\epsilon, m)$ returns the number of output spikes that result for an input ϵ given memory m. In the case of an ideal memory, k is solely a function of the input pattern ϵ.

For the general ASSAM with an ideal memory, if $x \in \{A, B\}$ and is x deviant, then the ESI is given by

$$\mathrm{ESI} = \frac{\sum_{\epsilon \in \mathcal{E}} k(\epsilon)\,[P(\epsilon \mid X = x) - P(\epsilon \mid X \neq x)]}{\sum_{\epsilon \in \mathcal{E}} k(\epsilon)\,[P(\epsilon \mid X = x) + P(\epsilon \mid X \neq x)]}, \tag{6.1}$$

where $\sum_{\epsilon \in \mathcal{E}}(\cdot)$ sums over all possible input patterns. The question now arises as to how one should choose the function $k(\epsilon)$ to maximize ESI.

First note that, from inspection of equation 6.1, $\forall \epsilon \in \mathcal{E}$ such that

$$P(\epsilon \mid X = x) = P(\epsilon \mid X \neq x),$$

changing $k(\epsilon)$ has no effect on the ESI because its contribution to the numerator is zero. Along similar lines, $\forall \epsilon \in \mathcal{E}$ such that $P(\epsilon \mid X = x) < P(\epsilon \mid X \neq x)$, setting $k(\epsilon) = 0$ maximizes ESI with respect to $k(\epsilon)$; $k(\epsilon) > 0$ will always lower the ESI. Informally, this means that emitting any spikes in response to an input pattern that is more likely to be generated by a standard than a deviant always leads to suboptimal ESI.

It now remains to set the $k(\epsilon)$, $\forall \epsilon \in \mathcal{E}$ such that $P(\epsilon \mid X = x) > P(\epsilon \mid X \neq x)$ in such a way that ESI is maximized. Observe that increasing $k(\epsilon)$ always moves ESI in the direction of

$$\frac{P(\epsilon \mid X = x) - P(\epsilon \mid X \neq x)}{P(\epsilon \mid X = x) + P(\epsilon \mid X \neq x)}$$

(see the appendix). Consequently, setting

$$k(\epsilon) = \begin{cases} K & \epsilon = \arg\max_{e} \left\{ \dfrac{P(e \mid X = x) - P(e \mid X \neq x)}{P(e \mid X = x) + P(e \mid X \neq x)} \right\} \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

($K > 0$) achieves the global maximum for the ESI. Choosing $k(\epsilon)$ in this way can be interpreted as configuring the abstract SSA model to generate spikes only when it receives the input pattern most likely to signify the deviant, thus avoiding mistaken classifications to the greatest degree possible.

A similar procedure can be followed to maximize the ESI for a general ASSAM with memory m in the case where x is the deviant: the function

$k(\epsilon, m)$ is set to K only for the ϵ and m that maximize

$$\frac{P(\epsilon, m \mid X = x) - P(\epsilon, m \mid X \neq x)}{P(\epsilon, m \mid X = x) + P(\epsilon, m \mid X \neq x)}$$

and set to zero elsewhere. Following this rule, the ASSAM generates spikes only when the input and memory together almost certainly indicate a deviant. This strategy we refer to as *hedging*, because it promotes inactivity in uncertain circumstances to secure a high performance.

The concomitant of hedging is that the expected spike count per tone, C, is lowered. This latter quantity is given by

$$E\{C\} = \sum_{\epsilon,m} k(\epsilon, m)\, P(\epsilon, m), \tag{6.3}$$

from which it is clear that the more $k(\epsilon, m)$ are lowered to improve the SI, the more the average spike count will fall. The key problems with hedging are twofold. First, there is no evidence that real neurons emit spikes only under very "safe" circumstances, for example, when a deviant is preceded by very many standards and the noisy neural input happens to heavily favor the deviant interpretation. If this were the case, one would have to measure very long sequences of tones until any spikes were generated, and, in fact, the experimental data show that neurons fire quite actively in response to oddball stimuli (Von der Behrens et al., 2009). Second, and more generally, a system that very rarely responds to any input, standard or deviant, communicates very little information.

This last point is worth reflecting on, as we have until now sought to maximize the SI value as the sole objective function. However, from the above considerations, it becomes clear that, for example, tuning the parameters of an SSA model to maximize SI will inevitably result in a model that is extremely insensitive. This trade-off was encounted in section 4.4 (see also Figure 6). Using a very slow recovery rate (β) on the adaptation meant that a very long—and hence unlikely—sequence of standards was required to obtain a deviant response. Consequently, such responses were very rare (lower spike count per tone) but more often correct (higher SI).

In closing this section, we consider how to maximize the SI in such a way that the unit remains responsive. The most natural starting point is to attempt to maximize the SI, subject to the constraint that the expected spike count per tone is held fixed at some suitable value, C_{avg}. This approach fails, however, because one can maximize the SI by hedging as before (e.g., using equation 6.2) and then set

$$K = \frac{C_{avg}}{P(\epsilon)}. \tag{6.4}$$

According to this solution, any response to a tone is still rare, but when it is elicited, the number of spikes is so large that the average spike count is kept artificially high. Of course, this solution is hardly an improvement on the original problem; nevertheless, it represents the global optimum that fulfills the stated requirements.

A better solution introduces an additional constraint: as well as fixing the expected spike count at C_{avg}, it states that the maximum response for any given tone cannot exceed $C_{\max}$. Maximizing the SI with these constraints is difficult; however, a similar but suboptimal solution, which fulfills these constraints and retains a tractable form, is an ASSAM that operates as follows. An input pattern is received across an infinite array of Poisson neurons packed with a density of $\rho = 50$ inputs per octave. Each input has $\sigma = 0.25$ octaves, $r_{\max} = 50$ spikes per second and $T = 10$ ms. In section 5.5, we used a Fisher information analysis to show that the gaussian approximation for the distribution of an unbiased estimator $\hat{f}$ of the tone frequency f is

$$p(\hat{f} \mid f) = \frac{1}{\sqrt{2\pi/J_\epsilon}} \exp\left[\frac{(\hat{f} - f)^2}{-2/J_\epsilon}\right], \tag{6.5}$$

and in this case, $J_\epsilon = 250.11$. Tones are presented at frequencies $f_A = 0$ and $f_B = \triangle f$, where $\triangle f = 0.15$ octaves.

The memory in this model is based on the depressing-standard criterion introduced in section 4.4, with $\alpha = 0.5$ and $\beta = 0.1$. This criterion states that if an incoming tone is classified as A, then a value proportional to the memory m_A is returned, and then m_A depresses while the memory m_B recovers. The opposite occurs if the incoming tone is classified as B.

A reduced ASSAM that takes action on the basis of the identification A or B cannot take into account the degree of certainty in the classification. For a general ASSAM that operates directly on the evidence ϵ, in this case, $\hat{f}$, we can modify the spiking rule as follows:

$$k(\hat{f}, m) = \begin{cases} C_{\max}\, m_A & \dfrac{P(\hat{f} \mid f = f_A)}{P(\hat{f} \mid f = f_B)} > \eta \\ C_{\max}\, m_B & \dfrac{P(\hat{f} \mid f = f_B)}{P(\hat{f} \mid f = f_A)} > \eta \\ 0 & \text{otherwise} \end{cases}. \tag{6.6}$$

Now the likelihood ratio of one tone to the other has to exceed a threshold, η, in order for spikes to be emitted. Letting $\eta = 1$ leads to a configuration equivalent to the reduced ASSAM. Using a large value of η means that the input pattern has to strongly indicate either A or B in order for the model to respond.

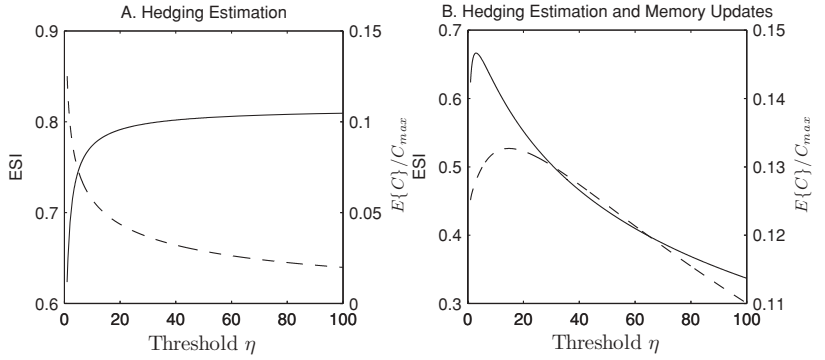

Figure 11: Expected ESI (solid lines) and normalized expected spike count (dashed lines) for (A) an algorithm that generates spikes only when the likelihood ratio is greater than η (cf. equation 6.6) and (B) an algorithm that generates spikes and causes the memory to adapt only when the likelihood ratio is greater than η. (Take note of the different scale on the left and right ordinate axes of the two plots.)

In our first evaluation of this function, spikes are returned according to rule 6.6, but memory updates take place normally. The results are plotted in Figure 11A. Consistent with our analysis, increasing η is shown to increase the ESI (solid line), as spikes are returned only for a sufficiently large likelihood ratio. At the same time, the average spike count (dashed line) decreases. Using a plot of this kind, one can set the sensitivity of the ASSAM by placing a lower limit on the spike count and finding the equivalent ESI at that point.

In this approach, the memory m_A depresses if the input pattern is identified as tone A—even if the probability of the input pattern given A only slightly exceeds that given B. Another modification of this algorithm is, like the spike rule, to allow the memory to depress only when the tone is identified with a given certainty. Thus, for tones $x, y \in \{A, B\}$, $x \neq y$, the memory update rule (i.e., $l(\epsilon, m)$) becomes

$$m_x = \begin{cases} \alpha m_x & \dfrac{P(\hat{f} \mid f = f_x)}{P(\hat{f} \mid f = f_y)} > \eta. \\ m_x + \beta(1 - m_x) \text{ otherwise.} \end{cases} \tag{6.7}$$

The result of using these spiking and update rules together with threshold parameter η is shown in Figure 11B.

7 Discussion

Stimulus-specific adaptation in biological neurons can be seen as a specific solution to the more general problem of highlighting rare events in a sequence. In this letter, we have characterized solutions to this general problem by means of abstract SSA models, which operate explicitly according to the known distribution of the input and some rudimentary memory scheme for storing those inputs (e.g., a buffer or an adaptation and recovery function).

Reduced ASSAMs map most naturally onto artificial neural networks that divide the labor between two (sets of) layers, the first of which improves the estimate of the tone location (e.g., a winner-take-all or lateral inhibition network) and the second of which serves as a memory (e.g., an array of depressing synapses). The N-spike criterion described in sections 4.2 and 5.1, in which units compete to accumulate Poisson input spikes up to a threshold, is analogous to a winner-take-all dynamic characterized by a pulse of excitation followed by a long-lasting wave of inhibition. Evidence consistent with such a mechanism is supplied by the SSA studies performed in subcortical regions where the responses of adapting units are phasic, being typically restricted to the first 50 ms of the tone or less (Anderson et al., 2009; Malmierca et al., 2009; Pérez-González et al., 2005).

Similar results have been obtained in cortex, although the pattern is less uniform. Von der Behrens et al. (2009) recorded the responses to 200 ms tones in rat primary auditory cortex and found that SSA was statistically significant only in the phasic portion, which lasted up to 50 ms. There was little activity in the remaining 150 ms; in fact, the poststimulus time histograms (PSTH) obtained from spontaneously active units showed a reduction below the spontaneous rate during this period, suggesting a slow inhibitory input. This conclusion is supported by current source density profiles measured in rat primary auditory cortex in response to oddball stimuli (Szymanski, Garcia-Lazaro, & Schnupp, 2009), which showed a current sink forming in cortical layers III–IV that lasted 20 to 40 ms, immediately followed by a current source throughout the remainder of the tone. Current sinks and current sources are thought to indicate, respectively, net excitatory and inhibitory currents in a small volume of tissue (Szymanski et al., 2009).

Contrary to this pattern, the mean PSTHs recorded by Ulanovsky et al. (2004, 2003) in cat auditory cortex consisted of a brief pulse followed by a tonic response for the duration of the tone, with the authors stating that the SSA response was more prominent in the sustained portion than the onset. Various factors could account for this difference: (1) the presentation rate—the experiments in IC used more rapid tone sequences than those carried out in cortex; (2) use and type of anesthetic—the results reported by Von der Behrens et al. (2009) were obtained from awake, behaving rats, while those reported by Ulanovsky et al. (2003, 2004) were conducted in anesthetized

cats; and (3) species—Ulanovsky et al. performed their experiments in cats, while all the other studies cited used rodents.

In the physiological experiments we referred to above, we commented on the pattern of excitation and inhibition in the response of the same neurons in which SSA was detected, whereas the reduced ASSAM implementing the N-spike criterion envisages this pattern arising in a separate layer whose output then propagates to another layer containing the SSA neurons. In order for stimulus-specific adaptation to arise in the final layer, we must posit a transformation between the input and output layers, which implements a form of memory. The transitory nature of this memory, as evidenced by the fact that it can relearn the stimulus statistics shortly after the standard and deviant tones are swapped (Ulanovsky et al., 2004), has led us in this letter to disregard forms of plasticity that effect more permanent changes in the synaptic connectivity, such as spike-time-dependent plasticity, in favor of those that bring about more short-lived changes, and synaptic depression in particular (Tsodyks & Markram, 1997).

The depressing-standard criterion presented in section 4.4 and elsewhere is based on the principle that the resources associated with the transmission of more common stimuli are more often depleted (called "fatigue models" by Grill-Spector et al., 2006). Several reseachers have proposed tonotopically ordered inputs feeding into strongly depressing (e.g., thalamocortical) synapses as an explanatory framework within which to interpret the data from SSA experiments (Von der Behrens et al., 2009; Szymanski et al., 2009; Ulanovsky et al., 2004, 2003; Ringo, 1996). This is broadly consistent with a reduced ASSAM that combines a tone estimation phase with the depressing standard criterion. In closing, we shall comment on existing results that such an ASSAM explains and results that an ASSAM predicts for other experimental designs. In terms of explanations of the data, we make the following remarks.

- *Discriminability principle*. Using a larger frequency difference ($\triangle f$) results in greater SI values. The ASSAM accounts for this result in one of two ways. For discrete tone classification (e.g., N-spike), the occasional misclassification of tones causes an average overlap in the uptake of synaptic resources. These errors are less frequent when the tones are easier to discriminate. A similar principle applies to models that activate a range of synapses according to, for example, a gaussian profile (see section 5.4): tones that are more widely separated consume more distinct resources.
- *Rarity principle*. Increasing the rarity of the deviant results in greater SI values. A lower value for p_{dev} implies a longer average interval between deviant tones, so that the synaptic depression associated with deviants has more time to recover, and conversely, the synaptic depression associated with standards has less time to recover.

- SSA should be principally attributable to depression in response to the standard rather than facilitation in response to the deviant. The conventional presentation of SSA results (PSTH for the standard, deviant and control conditions) tends to obscure this information. However, the timescales of adaptation reported by Ulanovsky et al. (2004, Figure 5C and 5D) supports the idea that both the standard and deviant responses adapt and that SSA is consequence of greater adaptation to standards (but see comments below).
- Similarly, inserting a silent period instead of a deviant tone produces no activity (Von der Behrens et al., 2009). The ASSAM does not explicitly predict a standard and then generate a deviant response when one fails to arrive; rather, the response is due to the deviant itself. Silence would not be expected to generate any activity above the spontaneous rate.
- *Locality principle*. The tones in a finite, recent history of the oddball sequence should determine what the deviant is. In the ASSAM, this feature is controlled by the choice of the α and β (reciprocal) time constants. If the standard and deviant tones are swapped, then the memory adjusts to reflect the new statistics after a period of approximately $1/\beta$ tones.[5] This is in line with the switching oddball paradigm used by Ulanovsky et al. (2004). Also, both Ulanovsky et al. (2004, Figure 7) and Von der Behrens et al. (2009, Figure 4) show that the response to a deviant depends on how many standards precede the deviant: a larger number of standards enhances the deviant response up to an asymptotic limit consistent with an exponential recovery.
- *Hyperacuity*. SSA is registered for frequency changes smaller than the bandwidth of the receptive field of the neuron (e.g., Ulanovsky et al., 2003). This will generally be the case for the ASSAM, as the SSA neuron effectively sums the output of an array of units with narrower tuning curves, which Miller, Escabí, Read, and Schreiner (2001) refers to as a "constructive inheritance" model of convergence. The presence of convergent input would seem to be supported by the fact that in many studies, the neurons that exhibit SSA are those with the widest receptive fields (Malmierca et al., 2009; Pérez-González et al., 2005, both in IC), although Ulanovsky et al. (2004) found no correlation between the width of the frequency response area and the SI for cells in auditory cortex.

[5]Of course, the memory employed in the depressing-standard criterion does not strictly satisfy the locality principle as the memory takes an infinite time to decay to zero, and so a single tone will still register an effect at every subsequent time. Nevertheless, we take its influence to be negligible after two time constants of recovery have elapsed.

A critical test of the model's veridicality concerns the response to deviant tones embedded in contexts other than an oddball sequence. A recent study (Nelken and Ta'aseh, 2009; Ta'aseh, Yaron, & Nelken, 2009) compared the response of single neurons to deviant tones presented in a normal oddball sequence, at the same rate without the standard tones, and in a "deviant among many standards" configuration, where rather than using a single standard frequency, the standard tones were distributed randomly over a wide frequency range. The average response to deviants in the oddball sequence was slightly lower than measurements taken when the standards were omitted—a result consistent with the ASSAM model if cross-frequency adaptation or tone misclassifications are assumed. Further experiments demonstrated that the mean response to deviants presented among many standards was similar to that recorded for deviants presented in a single-standard configuration. The investigators also developed an SSA model based on adaptation in bands, which predicted a smaller response to deviants in the oddball configuration. Nelken and Ta'aseh (2009) were thus led to reason that the bandwidths of the model could be modulated in a task-specific way, in particular, that they could become narrower for oddball sequences. At present, the parameters used in the deviant among many standards configuration have not been published in sufficient detail to assess the response of the ASSAM model.

We now consider the predictive power of an abstract SSA model. Prediction is envisaged as a two-stage process. In the first stage, the parameters σ, α, and β are adjusted so that the expected SIs obtained from the model match the SIs recorded in physiological experiments for various choices of $\triangle f$ and p_{dev}. The finer details of the model-fitting procedure we leave open at this stage. For example, one could also attempt to match $E\{D_A\}$, $E\{S_A\}$, and so on with d_A, s_A, and so on, respectively, in order to obtain a better model fit. Furthermore, any overall stimulus preference in a neuron (e.g., $s_A \gg s_B$) would also have to be accounted for by weighting the contribution of each depressing channel to the SSA neuron.

Having calibrated the ASSAM in this manner, the second stage would consist of supplying the model with "unseen" stimuli and comparing the output to measurements taken in the same cell used to calibrate the model. In the simplest instance, these could be oddball sequences with parameter pairs $(\triangle f, p_{dev})$ that were not used in stage 1. However, one could also subject the model to sequences generated by other kinds of random process, such as Markov chains or repeated frozen sequences, and then conduct an experiment to test the firing rates and SI values predicted by the model against those observed in the actual cell. An attractive feature of this model therefore is its potential to predict, or account for, a wide range of responses to complex stimuli in real neurons, while requiring the calibration of only a few, simple parameters such as α, β, and σ.

Knowing how successfully one can calibrate the ASSAM and then use it to make predictions will serve to validate or invalidate the assumptions

about physiological mechanisms engaged in the production of SSA (e.g., depression). More specifically, it may help to address the question of whether SSA should be labeled "reactive" or "predictive."[6] The ASSAM is reactive in the sense that it neither responds in the absence of input nor attempts to model the transition probabilities between tones. For this reason, while it is capable of implicitly capturing the distribution and entropy of its atomic inputs (e.g., tones A and B) (see section 5.3), it is incapable of learning patterns constructed from these atoms, whether stochastic or deterministic, still less detecting violations in those patterns. Mismatch negativity, on the other hand, is predictive, in the sense that it is elicited by violations in even quite sophisticated patterns and even by the absence of an expected stimulus (Yabe, Tervaniemi, Reinikainen, & Näätänen, 1997). Granted the premise that SSA is a reactive phenomenon, this letter provides a mathematical framework within which to assess the extent of the role played by SSA, in the production of mismatch negativity, to guide the design of computational models of SSA, and to evaluate the computational power of SSA.

Appendix

A.1 Frequency Separation. Two different measures of frequency separation are used in the literature cited in this article. Several authors (Anderson et al., 2009; Malmierca et al., 2009; Ulanovsky et al., 2004, 2003) use the normalized frequency difference:

$$\triangle f_{norm} \triangleq \frac{f_2 - f_1}{\sqrt{f_1 f_2}} \tag{A.1}$$

where f_1 and f_2 have Hertz units, while others use the octave frequency separation (e.g., Von der Behrens et al., 2009). The normalized frequency difference can be converted to a frequency separation measured in octaves ($\triangle f$) using

$$\triangle f \triangleq \log_2 \left(\frac{\triangle f_{norm}^2 + 2}{2} + \sqrt{\frac{(\triangle f_{norm}^2 + 2)^2}{4} - 1} \right). \tag{A.2}$$

A.2 *N*-Spike from Multiple Units. In section 5.1, the integral expression in equation 5.1 is approached by first rewriting the product of a sum as a

[6] We do not intend the labels "reactive" and "predictive" to introduce a strict dichotomy. For example, the ASSAMs discussed in this letter are reactive, but they also require a memory and can employ other forms of feedback, such as global or lateral inhibition.

nested sum of products and moving the integral inside:

$$
\begin{aligned}
&P(\hat{X} = X_j \mid X = X_k) \\
&= \int_0^\infty \frac{r_j^N}{(N-1)!} t^{N-1} \exp\left(-r_j t\right) \prod_{i \neq j}^{M} \sum_{n=0}^{N-1} \frac{(r_i t)^n}{n!} \exp(-r_i t)\, dt \\
&= \int_0^\infty \frac{r_j^N}{(N-1)!} t^{N-1} \exp\left(-r_j t\right) \times \\
&\quad \left[\sum_{n_1=0}^{N-1} \cdots \sum_{n_{j-1}=0}^{N-1} \sum_{n_{j+1}=0}^{N-1} \cdots \sum_{n_M=0}^{N-1} \prod_{i \neq j}^{M} \frac{(r_i t)^{n_i}}{n_i!} \exp(-r_i t) \right] dt \\
&= \sum_{n_1=0}^{N-1} \cdots \sum_{n_{j-1}=0}^{N-1} \sum_{n_{j+1}=0}^{N-1} \cdots \sum_{n_M=0}^{N-1} \frac{r_j^N}{(N-1)!} \frac{\prod_{i \neq j}^{M} r_i^{n_i}}{\prod_{i \neq j}^{M} n_i!} \\
&\quad \int_0^\infty t^{(N-1+\sum_{i \neq j}^{M} n_i)} \exp\left(- \sum_{i}^{M} r_i t \right) dt.
\end{aligned}
$$

Then, noting that $\int_0^\infty t^{k-1} \exp(-\lambda t) dt = \lambda^{-k}(k-1)!$ one can solve the integral and rearrange to obtain the closed-form expression

$$
\sum_{n_1=0}^{N-1} \cdots \sum_{n_{j-1}=0}^{N-1} \sum_{n_{j+1}=0}^{N-1} \cdots \sum_{n_M=0}^{N-1} \frac{\left(N + \sum_{i \neq j}^{M} n_i - 1\right)!\, r_j^N \prod_{i \neq j}^{M} r_i^{n_i}}{(N-1)! \prod_{i \neq j}^{M} n_i! \left(\sum_i^M r_i\right)^{N + \sum_{i \neq j}^{M} n_i}}. \tag{A.3}
$$

A.3 Explicitly Maximizing ESI. This proof relates to section 6, in which it is stated that increasing $k(\epsilon)$ moves the ESI in the direction of $[P(\epsilon \mid X) - P(\epsilon \mid Y)] / [P(\epsilon \mid X) + P(\epsilon \mid Y)]$. First, using the abbreviations

$$
\begin{aligned}
\phi(\epsilon) &\equiv P(\epsilon \mid X) - P(\epsilon \mid Y) \\
\chi(\epsilon) &\equiv P(\epsilon \mid X) + P(\epsilon \mid Y),
\end{aligned}
$$

one can write the derivative of the ESI as defined in equation 6.1 as

$$
\begin{aligned}
\frac{\partial\, \mathrm{ESI}}{\partial k(\iota)} &= \frac{\partial}{\partial k(\iota)} \left\{ \frac{\sum_{\epsilon \in \mathcal{E}} k(\epsilon) \phi(\epsilon)}{\sum_{\epsilon \in \mathcal{E}} k(\epsilon) \chi(\epsilon)} \right\} \\
&= \frac{\phi(\iota) \sum_{\epsilon \in \mathcal{E}} k(\epsilon) \chi(\epsilon) - \chi(\iota) \sum_{\epsilon \in \mathcal{E}} k(\epsilon) \phi(\epsilon)}{\left[\sum_{\epsilon \in \mathcal{E}} k(\epsilon) \chi(\epsilon)\right]^2} \\
&= \frac{\phi(\iota)}{\sum_{\epsilon \in \mathcal{E}} k(\epsilon) \chi(\epsilon)} - \frac{\chi(\iota)\, \mathrm{ESI}}{\sum_{\epsilon \in \mathcal{E}} k(\epsilon) \chi(\epsilon)},
\end{aligned} \tag{A.4}
$$

from which it follows that

$$\mathrm{ESI} < \frac{\phi(\iota)}{\chi(\iota)} \Leftrightarrow \frac{\partial\,\mathrm{ESI}}{\partial k(\iota)} > 0. \tag{A.5}$$

From equation A.5, one may infer that increasing $k(\iota)$ always moves the ESI in the direction of $\phi(\iota)/\chi(\iota)$, that is, $[P(\epsilon \mid X) - P(\epsilon \mid Y)]/[P(\epsilon \mid X) + P(\epsilon \mid Y)]$.

Acknowledgments

This work was supported by the European Community's Seventh Framework Programme (grant no. 231168—SCANDLE). We thank István Winkler for his helpful comments on the manuscript.

References

Anderson, L. A., Christianson, G. B., & Linden, J. F. (2009). Stimulus-specific adaptation occurs in the auditory thalamus. *Journal of Neuroscience, 29*, 7359–7363.

Arbib, M. A. (1995). *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.

Brunel, N., & Nadal, J.-P. (1998). Mutual information, Fisher information, and population coding. *Neural Computation, 10*, 1731–1757.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*, 14–23.

Koch, C. (1998). *Biophysics of computation*. New York: Oxford University Press.

Liberman, M. C. (1978). Auditory-nerve responses from cats raised in a low-noise chamber. *J. Acoust. Soc. Am., 63*, 442–455.

Malmierca, M. S., Cristaudo, S., Pérez-González, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anaesthetized rat. *Journal of Neuroscience, 29*, 5483–5493.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13*, 1257–1269.

Miller, L., Escabí M., Read, H., & Schreiner, C. (2001). Functional convergence of response properties in the auditory thalamocortical system. *Neuron, 32*, 151–160.

Nelken, I., & Ta'aseh, N. (2009). The coding of surprise in auditory cortex of rats. *Frontiers in Human Neuroscience. Conference Abstract: MMN 09 Fifth Conference on Mismatch Negativity (MMN) and its Clinical and Scientific Applications.* Available online at http://www.frontiersin.org/CommunityAbstract Details.aspx?ABS_DOI=10.3389/conf.neuro.09.2009.05.66.

Nelken, I., & Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation. *Journal of Psychophysiology, 21*, 214–223.

Nicholls, J. G., Martin, A. R., Wallace, B. G., & Fuchs, P. A. (2001). *From neuron to brain* (4th ed.). Sunderland, MA: Sinauer Associates.

Oster, M., Douglas, R., & Liu, S.-C. (2009). Computation with spikes in a winner-take-all network. *Neural Computation, 21*, 2437–2465.

Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables and stochastic processes* (4th ed.). New York: McGraw-Hill.

Pérez-González, D., Malmierca, M. S., & Covey, E. (2005). Novelty detector neurons in the mammalian auditory midbrain. *European Journal of Neuroscience, 22*, 2879–2885.

Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavorial Brain Research, 76*, 191–197.

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. USA, 90*, 10749–10753.

Shamir, M. (2009). The temporal winner-take-all readout. *PLoS Computational Biology, 5*, e1000286.

Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal, 5*, 173–194.

Szymanski, F. D., Garcia-Lazaro, J. A., & Schnupp, J. W. H. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex. *Journal of Neurophysiology, 102*, 1483–1490.

Ta'aseh, N., Yaron., A., & Nelken, I. (2009). Specific deviance detection in rat auditory cortex. *Program No. 163.15/T13 2009 Neuroscience Meeting Planner*. Chicago: Society for Neuroscience.

Tsodyks, M. V., & Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. USA, 94*, 719–723.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience, 24*, 10440–10453.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience, 6*, 391–398.

Von der Behrens, W., Bäuerle, P., Kössl, M., & Gaese, B. H. (2009). Correlating stimulus-specific adaptation of cortical neurons and local field potentials in the awake rat. *Journal of Neuroscience, 29*, 13837–13849.

Winkler, I., Denham, S. L., & Nelken I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitives Sciences, 13*, 532–540.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport, 8*, 1971–1974.

Received February 17, 2010; accepted July 27, 2010.