HAL
open science

Contextual modulation of auditory responses predicted by statistical inference

Brian Fischer, Sophie Deneve

To cite this version:

Brian Fischer, Sophie Deneve. Contextual modulation of auditory responses predicted by statistical inference. Cinquième conférence plénière française de Neurosciences Computationnelles, "Neurocomp'10", Aug 2010, Lyon, France. hal-00553432

HAL Id: hal-00553432

https://hal.science/hal-00553432v1

Submitted on 16 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONTEXTUAL MODULATION OF AUDITORY RESPONSES PREDICTED BY STATISTICAL INFERENCE

Brian J Fischer, Sophie Deneve
Group for Neural Theory, Ecole Normale Superieure
29 rue d'Ulm, 75005 Paris
France
brian.fischer@ens.fr, sophie.deneve@ens.fr

ABSTRACT
Auditory neurons exhibit complex nonlinear and context-dependent responses that are difficult to capture by standard modeling techniques. Rather than using a bottom up approach and characterizing the responses of individual cells as a function of their input, we propose to use a top-down, normative approach to analyzing auditory processing. We develop a model of spiking neurons that perform probabilistic inference to estimate the state of auditory environment from sensory signals. The model predicts a form for the nonlinear, context-dependent modulation of inputs to central auditory neurons. We show that a simple model based on auditory inference can explain the presence of multiple aspects of contextual modulation in both frequency and time that are observed in the auditory system.

KEY WORDS
Audition, inference, contextual modulation, nonlinearity

1. Introduction

Our understanding of sensory representations in the brain has been dominated by the idea of the neural receptive field. This concept is best characterized by Hubel and Wiesel's [1] classic description of the responses of visual cortical neurons. Here, a neuron responds to a visual stimulus by applying a weight to each region of the visual field and summing the responses. In the extension of this approach, neurons are characterized by a fixed selectivity for stimulus features such as the orientation of a bar of light or the frequency of a tone, and sensory representations are characterized by manner in which preferred selectivity for stimulus features varies over the population. Either explicitly, or implicitly, sensory systems are often thought of as applying a fixed set of filters to the sensory input.

In the auditory system, the stimulus selectivity of a neuron is usually defined by a spectrotemporal receptive field (STRF). The STRF is the time-frequency filter that, when applied to a time-frequency representation of the stimulus, produces the best fit to the neural response [2]. The accuracy of the STRF model is limited, however, because of the presence of nonlinearities in many neural responses [3]. The consequence of such nonlinearities is that models fit to responses obtained using one type of stimulus often do not generalize to other sets of stimuli [4]. Second, receptive field properties of cortical neurons are subject to significant variability with context ([5], [6]). Again, this aspect of cortical responses causes models fit from one data set to fail to generalize to other contexts. An alternative approach to characterizing sensory representations is to derive the optimal neural representation given a theoretical goal of sensory processing.

In light of the growing evidence that many aspects of perception are consistent with optimal sensory processing [7], a promising functional approach is to construct models of neural circuits that perform optimal inference. In particular, the tasks humans face in natural situations require the combination of multiple noisy and ambiguous sensory cues, as well as the use of prior knowledge, either innate or acquired from previous experiences. Such incomplete and imperfect sensory cues and prior knowledge can only provide probabilistic information. With this view, the goal of sensory processing is to perform optimal statistical inference. Several models have been proposed to explain the neural basis of optimal sensory perception ([8],[9],[10],[11],[12],[13]). However, such models have not addressed the nonlinear processing that occurs in auditory inference. Here we show that a simple model based on auditory inference exhibits multiple aspects of nonlinear, contextual modulation that are seen in the auditory system.

2. Auditory environment

The task of the auditory system is to infer the structure of the auditory environment from the sensory signals that are received at the ears. From a functional point of view, we describe the auditory system as doing probabilistic inference using a model of the world. The structure of this model is given by a factorial hidden Markov model (FHMM, [10],[14]; Figure 1). The FHMM consists of a set of independent sources, each evolving according to its own dynamics, that combine to produce the responses of auditory neurons. The essence of this model is the supposition that the auditory system is attempting to explain the sensory

data as the result of a set of independent causes. This model therefore shares the basic features of independent component analysis-based models that have been proposed to explain properties of the auditory [15] and visual [16] systems.

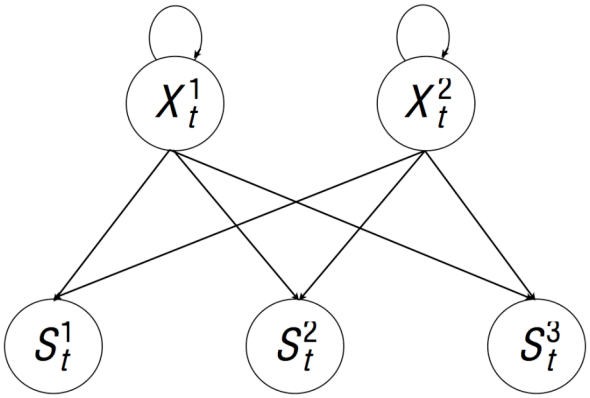

Figure 1: The causal model that describes the auditory environment is given by a factorial hidden Markov model. The environment consists of a number of independent sources that combine to produce spiking responses in sensory neurons.

The FHMM is characterized by the dynamics of the underlying sources and how the sources influence the sensory neurons. We assume that each source is a binary variable that evolves with first-order Markov dynamics. The dynamics of the ith source $X_t^i \in \{0,1\}$ is therefore characterized by two transition rates $r_{on}^i = P(X_{t+dt}^i = 1 \mid X_t^i = 0)/dt$ and $r_{off}^i = P(X_{t+dt}^i = 0 \mid X_t^i = 1)/dt$.

The sensory neurons are spiking neurons that encode the state of the auditory environment. The probability of spiking for each sensory neuron is modeled by a noisy-or function $P(S_t^j = 1 \mid \mathbf{X}_t) = 1 - (1 - dtq_{0j})\prod_{i=1}^{N}(1 - dtq_{ij})$, where N is the number of sources in the environment. The influence of the ith source on the spiking probability of the jth sensory neuron is described by the parameter q_{ij}. The set of these parameters over all sensory neurons for a given source is termed the causal field of the source. Given that the independent components of natural sounds correspond to bandpass filters [15], we will assume that the causal fields correspond to bands of frequency.

3. Auditory inference by spiking neurons

3.1 Inference equation

The inference problem is to determine the state of the environment from the spike trains produced by the sensory neurons. Ideally, one would compute the posterior distribution over the state of the environment given the history of spike trains, $P(\mathbf{X}_t \mid \mathbf{S}_{0:t})$. However, exact inference of this form for a large number of sources and sensory neurons is computationally intractable [14]. Therefore, we make an approximation that allows for nearly exact inference without the need to represent the joint probability distribution over a high dimensional space of sources.

Following [10], we propose that inference in the FHMM is done using the probabilities of each source given the history of sensory spike trains and the approximate state of the other sources in the environment. That is, inference is based on the probabilities $P(X_t^i \mid S_{0:t}, E[X_t^{j\neq i}])$ where $E[X_t^i] = P(X_t^i = 1 \mid S_{0:t}, E[X_t^{j\neq i}])$. This probability is computed by developing a recursive equation for the log-odds

$$L_t = \ln\left(\frac{P(X_t^i = 1 \mid S_{0:t}, E[X_t^{j\neq i}])}{P(X_t^i = 0 \mid S_{0:t}, E[X_t^{j\neq i}])}\right).$$

In the limit of small time steps, we find that the log-odds evolves according to the differential equation

$$\frac{d}{dt}L_t^i = \Phi(L_t^i) + \sum_{j=1}^{M}\omega_{ij} S_t^j$$

where the weights depend on the context. Specifically, the weights are given by

$$\omega_{ij} = \ln\left(1 + \frac{q_{ij}}{q_{oj} + \sum_{k\neq i} q_{kj} E[X_t^k]}\right)$$

so that the weights are suppressed when other causes explain the activity of the sensory neuron. This is a particular form of divisive inhibition that arises from the inference algorithm.

3.2 Inference with spiking neurons

The dynamics of inference can be mapped onto the dynamics of a population of spiking neurons ([8],[10]). Briefly, the membrane potential of the neuron that encodes the log-odds of the ith source is given by

$$V_t^i = L_t^i - G_t^i$$

where

$$\frac{d}{dt}G_t^i = \Phi(G_t^i) + O_t$$

and O_t is equal to one if the neuron fires at time t and is zero otherwise. The membrane potential of the neuron, therefore, is a function of the spikes from the sensory neurons and its own spike history. The neuron fires a spike when the membrane potential reaches a threshold.

The term G provides an estimate of the log-odds of the underlying source given the output spikes of the neuron. The stimulus is estimated from the output spike trains as

$$E[X_t^i] = \exp(G_t^i)/(1 - \exp(G_t^i)).$$

4. Simulation results

4.1 Contextual modulation of weights

We first illustrate that the neurons performing inference in the model will integrate spikes from the sensory neurons using context-dependent weights. For this example, we use a simple model environment with two sources and three sensory neurons. The two sources have overlapping causal fields that correspond to the overlap in frequency of two bandpass sources.

The system is able to accurately infer the states of the two sources from the spike trains produced by the sensory neurons (Figure 2). The inference is accurate even when both sources are on simultaneously. As seen in Figure 2, the weights applied to the sensory spikes are modified by the context. When the probability of source 1 being active is high, sensory information is weighted less heavily by the neuron representing the probability of source 2. In this system, neurons performing inference do not integrate their inputs with fixed weights, but rather modify their integrative properties by contextual information.

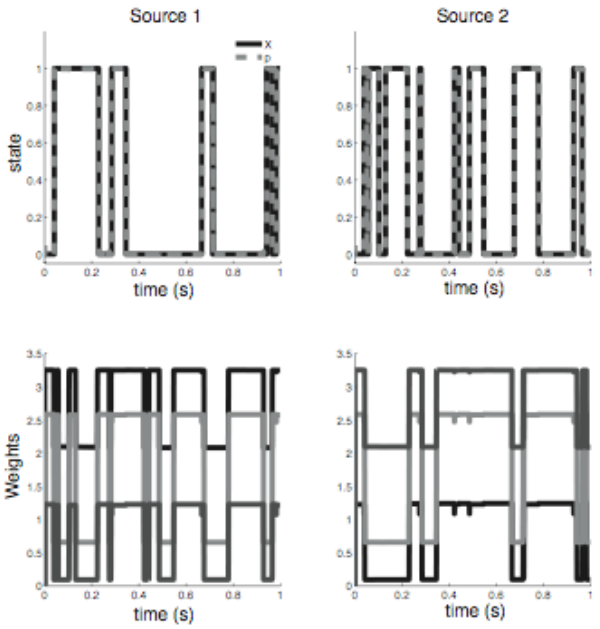

Figure 2: Modulation of weights by context. Top: The state of the environment (solid) and the estimate from the spiking neurons (dashed). Bottom: The weights applied to the sensory spikes to infer the state of source 1 (left) and source 2 (right). Note the decrease in weight due to the presence of the other source.

4.2 Two-tone rate suppression

A well known nonlinear property of many auditory neurons is the suppression of responses to preferred tonal stimuli by the presence of additional tones around the preferred frequency [17]. To examine two-tone rate suppression in the model, we considered a model environment consisting of 12 sources and 15 sensory receptors. The sources have overlapping causal fields, as in the previous example, corresponding to overlapping bandpass sources.

We found that neurons performing inference in the system showed nonlinear responses consistent with two-tone rate suppression (Figure 3). Given the form of the weights in the model, each neuron will have an excitatory or zero response to the presence of a single source in the environment. However, when the preferred stimulus was paired with stimuli that have overlapping causal fields, the response of the neuron was reduced relative to the response to the preferred stimulus alone (Figure 3).

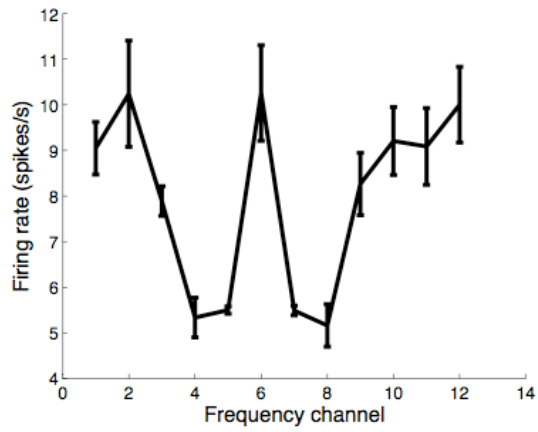

Figure 3: Two-tone rate suppression in a model neuron. The response at channel six is the response to source six alone. The other points are the responses to the simultaneous presence of source six and the given source. Error bars represent the standard deviation.

4.3 Forward suppression

Auditory neurons show modification of responses to temporal context. An example of temporal context dependence is forward suppression, which is seen in many auditory neurons [18]. Here, the occurrence of an input stimulus will cause a decrease in the response to subsequent inputs. This physiological response is thought to be related to the psychophysical phenomenon of forward masking [18]. To examine forward suppression in the model, we simulate a paired pulse paradigm [18].

We found that neurons performing inference in the model showed forward suppression (Figure 4). The response to a second pulse in a paired pulse stimulus was decreased relative to the response to the first pulse. The amount of suppression was greatest for the shortest interpulse intervals, as seen experimentally [18].

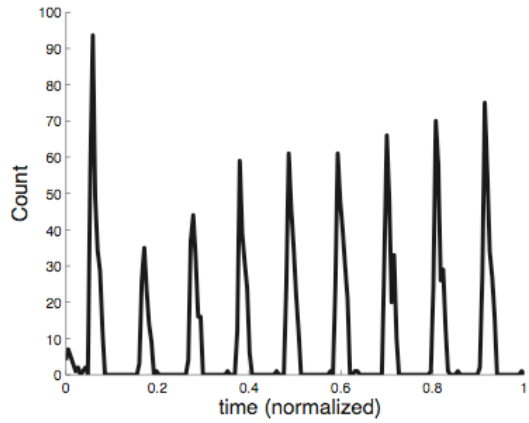

Figure 4: Response of a model neuron to pairs of pulses.

3. Conclusion

We propose that the nonlinear, context-dependent responses seen in the auditory system may be signatures of optimal inference. Rather than using a bottom up approach and characterizing the responses of individual cells as a function of their input, we used a top-down, normative approach to analyzing auditory processing. In this approach, neural responses provide a subjective, probabilistic interpretation of the auditory scene as the events that best predict the auditory input, rather than a deterministic function of this input. Using simple examples of this model structure, we are able to see several of the basic context dependent properties of auditory neurons.

What distinguishes this approach from approaches based on receptive fields is the fact that interpreting the sensory scene and consequent neural responses is a collective, cooperative process rather than a set of cells applying each a different fixed filter (i.e. spectrotemporal receptive field) to the input. Interpretation of the auditory input channels depend on the context, e.g. which features are currently and were previously present in the scene. Indeed, the auditory input is highly ambiguous, since different sounds will affect highly overlapping responses in the auditory nerve. The problem of source separation is to recover the independent sources from this highly redundant and clustered input. This model illustrates how a system designed to perform this type of inference can exhibit many of the properties seen in the auditory system.

The current form of the model has not addressed the problem of learning the statistics of the environment in an online manner. Such an approach is necessary to predict how sensory representations will adapt to changes in the statistics of the input. It may be possible to apply an algorithm based on online expectation maximization [19] to incorporate learning into the system. The incorporation of a learning rule into the system will add another mechanism for modifying responses by context. We suggest that a model of auditory processing based on online inference and learning will provide a framework for understanding the complex representation of the auditory environment by central circuits.

References

[1] D.H. Hubel & T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol. (Lond.) 160*, 1962, 106-154.

[2] A. Aertsen & P. Johannesma, The spectro-temporal receptive field: a functional characteristic of auditory neurons, *Biol. Cybern. 42*, 1981, 133-143.

[3] M. Ahrens, J.F. Linden, & M. Sahani, Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J. Neurosci. 28*, 2008, 1929-1942.

[4] G.B. Christianson, M. Sahani, & J.F. Linden, The consequences of response nonlinearites for interpretation of spectrotemporal receptive fields, *J. Neurosci. 28,* 2008, 446-455.

[5] N.M. Weinberger, Dynamic regulation of receptive fields and maps in the adult sensory cortex, *Annu. Rev. Neurosci. 18*, 1995, 129-158.

[6] J. Fritz, M. Elhilali, & S. Shamma, Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex, *Hearing Res. 206*, 2005, 159-176.

[7] D.C. Knill & A. Pouget, The Bayesian brain: The role of uncertainty in neural coding and computation, *Trends Neurosci*. *27*(12), 2004, 712-719.

[8] S. Deneve, Bayesian spiking neurons I: Inference, *Neural Computation 20*, 2008, 91-117.

[9] S. Deneve, Bayesian spiking neurons II: Learning, *Neural Computation 20*, 2008, 118-145.

[10] S. Deneve, T. Lochmann, & U. Ernst, Spike based inference in a network with divisive inhibition, *Proc. Neurocomp08*, Marseille, France, 2008.

[11] W.J. Ma, J. Beck, P. Latham, & A. Pouget, Bayesian inference with probabilistic population codes, *Nature Neurosci., 9*(11), 2006, 1432-1438.

[12] R. Rao, Bayesian computation in recurrent neural circuits, *Neural Computation 16*(1), 2003, 1-38.

[13] C. Eliasmith & C.H. Anderson, *Neural engineering* (Cambridge, MA: MIT Press, 2003).

[14] Z. Ghahramani & M. Jordan, Factorial hidden Markov models, *Machine Learning 29*, 1997, 245-273.

[15] M. Lewicki, Efficient coding of natural sounds, *Nature Neurosci 5*, 2002, 356-363.

[16] B.A. Olshausen & D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature 381*, 1996, 607-609.

[17] B. Delgutte, Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level, *Hearing Research 49*, 1990, 225-246.

[18] M. Wehr & A. Zador, Synaptic mechanisms of forward suppression in rat auditory cortex, *Neuron 47*, 2005, 437-445.

[19] G. Mongillo & S. Deneve, Online learning with hidden Markov models, *Neural Computation 20*, 2008, 1706-1716.