Predicting spike timing in highly synchronous auditory neurons at different sound levels

Bertrand Fontaine[1,2,3,4], Victor Benichoux[1,2], Philip, X. Joris [3], Romain Brette[1,2]

[1] Laboratoire Psychologie de la Perception, CNRS, Université Paris Descartes, Paris, France
[2] Equipe Audition, Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France
[3] Laboratory of Auditory Neurophysiology, University of Leuven, Leuven, Belgium
[4] Dominick P. Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, USA

Corresponding author: Romain Brette
Equipe Audition
Département d'Etudes Cognitives
Ecole Normale Supérieure
29, rue d'Ulm
75230 Paris Cedex 05
Tel: +33 (0)1.44.32.26.13
Email: romain.brette@ens.fr

Abstract

A challenge for sensory systems is to encode natural signals that vary in amplitude by orders of magnitude. The spike trains of neurons in the auditory system must represent the fine temporal structure of sounds despite a tremendous variation in sound level in natural environments. It has been shown in vitro that the transformation from dynamic signals into precise spike trains can be accurately captured by simple integrate-and-fire models. In this work, we show that the in vivo responses of cochlear nucleus bushy cells to sounds across a wide range of levels can be precisely predicted by deterministic integrate-and-fire models with adaptive spike threshold. Our model can predict both the spike timings and the firing rate in response to novel sounds, across a large input level range. A noisy version of the model accounts for the statistical structure of spike trains, including the reliability and temporal precision of responses. Spike threshold adaptation was critical to ensure that predictions remain accurate at different levels. These results confirm that simple integrate-and-fire models provide an accurate phenomenological account of spike train statistics, and emphasize the functional relevance of spike threshold adaptation.

Introduction

To localize sound sources in the horizontal plane, mammals rely mainly on interaural time differences (ITDs) at low frequencies. In cats, ITDs are smaller than 400 µs (Tollin and Koka, 2009) and behaviorally just noticeable differences in ITD can be as small as 20 µs (Wakeford and Robinson, 1974). The auditory system displays a number of specializations that reflect the required precision of fine temporal processing (Oertel, 1999; Trussell, 1999; Yin, 2002). In particular, in the cochlear nucleus (CN), low frequency bushy cells respond to acoustic signals from the ipsilateral ear with submillisecond precision (Joris et al., 1994; Louage et al., 2005; Joris and Smith, 2008). They project from both sides to binaural neurons in the medial superior olive (MSO), which respond to coincident input spikes, making them sensitive to ITDs (Yin and Chan, 1990). One challenge faced by this system is to encode ITD over the enormous range of stimulus intensities that the animals experience. Little data are available from MSO neurons, but responses from its targets (particularly the inferior colliculus : (Yin et al., 1986), Fig. 3) suggest that ITD tuning is surprisingly invariant to sound level. In response to tones, the response rate and temporal coding in bushy cells is less sensitive to sound level than in the auditory nerve (AN) (Joris et al., 1994, Recio-Spinoso, 2012). This also appears to be the case in response to noise (discussed as "compression" in van der Heijden and Joris, 2009). For example, bushy cells afford lower just noticeable differences for ITD discrimination, over a wider range of SPLs, than AN fibers (van der Heijden et al., 2011).

As reported in many sensory pathways, neurons adapt to input statistics (Brenner et al., 2000; Fairhall et al., 2001; Hosoya et al., 2005; Nagel and Doupe, 2006). This adaptation has been mostly described in terms of firing rate. ITD processing is original in that adaptation is found in the temporal responses of neurons. While previous works reported the effect of input level on spike jitter and reliability in cochlear nucleus (e.g. Louage et al., 2005), here we analyze and model the effect of input level on absolute timing (Michelet et al., 2012) in bushy cells. In order to describe the transformation of a continuous acoustical signal into a sequence of precisely timed spikes, we design a phenomenological model of CN responses that can predict every spike at different input levels, with a single set of parameters.

In vitro, several groups have shown that it is possible to accurately predict the precise time of spikes produced by a neuron in response to time-varying currents injected at the soma, using simple integrate-and-fire models (IF) (Jolivet et al., 2008; Gerstner and Naud, 2009; Rossant et al., 2010, 2011). In this paper we apply the same method to our CN *in vivo* single-unit recordings and find that simple IF models cannot predict the responses because they are too sensitive to level. We ask whether the addition of an adaptive threshold to our model could improve prediction. Spike threshold – the membrane voltage above which a spike is triggered - varies and depends on spike history (Azouz and Gray, 2000; Wilent and Contreras, 2005; Chacron et al., 2007). In vitro, the addition of a dynamic threshold to a simple IF model has been shown to improve the prediction of cell responses to injected random currents (Jolivet et al., 2008; Kobayashi et al., 2009; Rossant et al., 2010). In vivo, IF models with dynamical threshold can successfully reproduce experimental data in visual (Keat et al., 2001), electrosensory (Savard et al., 2011) and vestibular (Sadeghi et al., 2007) neurons. However, in the present study this approach is not sufficient to predict spikes when the input level is varied. We show that a threshold model with multiplicative spike-triggered adaptation (Brette, 2012) can accurately predict the timing of spikes in response to acoustical inputs across a broad range of levels. While the present work emphasizes the relevance of spike threshold adaptation, it also provides a

predictive model of bushy cell responses with very few parameters that can be used in studies of the sound localization pathway at the systems level, e.g. as inputs to binaural coincidence detectors.

Material and methods

The data presented in this paper represent a subset of the data collected in (Louage et al., 2005, 2006). All procedures were approved by the KU Leuven Ethics Committee for Animal Experiments and were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. The experimental methods are described in detail there and are only briefly summarized here. Pentobarbital anesthetized cats were placed in a soundproof room. A sealed acoustic driver was inserted into one or both exposed ear canals and calibrated with a 1/2 inch condenser microphone and a probe tube close to the eardrum. The trapezoid body (TB) was exposed via a ventral approach to the skull base. All data were recorded with glass micropipettes filled with 3 M NaCl. The neural signal was converted to spike times referenced to the stimulus onset with a peak detection triggering circuit with an accuracy of 1 μs.

Stimuli and data collection

The search stimulus was a binaural noise burst (duration 300 ms, repeated every 500 ms, 70 dB SPL, bandwidth 40 kHz). When the activity of a single fiber was isolated, the excitatory ear was determined. For each fiber encountered, a threshold tuning curve was obtained with a tracking algorithm that provided spontaneous rate, characteristic frequency (CF), and threshold. Short tone bursts at CF (duration 25 ms, repeated every 100 ms, 200 repetitions, rise-fall time 2.5 ms, starting in sine phase) were then presented at increasing sound pressure level (SPL) in 10-dB steps. Next, a rate-level function was obtained to a broadband Gaussian noise (1 s), repeated every 1.2 s , five to ten repetitions).

After a fiber's basic physiological parameters and rate-level functions were collected, a 1-second broadband noise (0.1-30kHz) with many repetitions was delivered, usually 35 to 100, to collect at least 3,000 spikes. In some cases, subsequently a second, independent 1-second noise token was similarly delivered. The first input level (overall level re 20 μPa) tested was usually 70 dB SPL, the next levels were usually 50, 30, 80, 60, and 10 dB SPL. Because the time we could record from a TB fiber was limited, for certain fibers not all levels were presented.

Data selection

Only a subpopulation of the available recordings was used in our analysis. Fibers of the TB were classified into different categories based on the shape of their post-stimulus time histogram (PSTH) (binwidth 0.1 ms) to short pure tone bursts at CF at different stimulus levels (Louage et al., 2005, 2006). We restrict our analysis to low-frequency fibers that show a phase-locked PSTH ("PHL") and which furthermore show the so-called "high-sync" property (Joris et al., 1994). The exact selection criteria and resulting database are stated at the beginning of the Results section. It was extensively discussed in previous papers (Joris et al., 1994; Louage et al., 2005; Joris and Smith, 2008) that the vast majority, if not all, of these "high-sync" TB fibers are axons of the two variety of bushy cells: both spherical and globular bushy cells (SBCs and GBCs, respectively).

Nevertheless, throughout the Results section we use the neutral term "TB fiber" to acknowledge the fact that the anatomical identity is not known with certainty for any given fiber.

Correlation analysis

In order to assess the synchronization properties of a neuron's response to different presentations of the same noise token at a single stimulus level, we construct the shuffled auto-correlogram (SAC) (Louage et al., 2005; Joris et al., 2006). Every possible trial pair is compared (except comparisons of a trial to itself, Fig. 2A1): time intervals between all spikes of the first train and all spikes of the second spike train are measured and tallied in a histogram (Fig. 2A2). Since SACs are symmetric, only forward time intervals are considered. The resulting histogram is then mirrored, yielding the SAC (Fig. 2A3). The SAC ordinate is normalized by $n(n-1)\Delta\tau r^2 D$ This factor eliminates the effect of average rate r, number of presentations n, choice of bin width $\Delta\tau$, and stimulus duration D. This scaling yields dimensionless bin values. The maximal value of the SAC is referred to as the correlation index (CI, Fig. 2A3). Uncorrelated spike trains result in a value of 1. A measure of the temporal precision is derived from the SAC by taking the width of the main lobe where the values are half of the SAC peak. We refer to this measure, given in ms, as half-height width (HHW, Fig. 2A3).

We use the same concept to analyze the effects of stimulus level on ongoing timing in TB as in (Michelet et al., 2012). Cross stimulus auto-correlograms (XAC) are computed between the responses of the same cell at two different stimulus levels. All possible pairs of trials between the two levels are taken into account (Fig. 2B1). As XACs are not symmetric, both forward and backward time intervals are tallied (Fig. 2B2, B3). The XAC ordinate is normalized by $n_1 n_2 \Delta\tau r_1 r_2 D$, where n_i and r_i are the number of presentations and firing rate of the *i*th response. If the responses to the lower level lead the responses to the higher level, the correlogram peak will be shifted to the left . The lag is defined as the position of the main lobe peak (Fig. 2B3).

Peripheral model

The model chain, describing the mapping from sounds to spike trains in cochlear nucleus bushy cells, is shown in Fig. 5A. The first element, shared by all the models we consider, is linear filtering. It summarizes the linear filtering properties of the afferents to the cochlear nucleus and of the neuron itself. It is characterized by an impulse response (Fig. 5A, auditory filter).

This impulse response is calculated by reverse correlation (RevCor). For a broadband noise stimulus, the RevCor filter $\mathrm{h(t)}$ is the average stimulus that elicits spikes (de Boer and de Jongh, 1978; Schwartz et al., 2006), that is, $\mathrm{h(t)} = \frac{1}{\mathrm{N}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \vec{s}(\mathrm{t_n})$, where $\mathrm{t_n}$ is the time of the $\mathrm{n^{th}}$ spike, $\vec{s}(\mathrm{t_n})$ is a vector containing the stimulus present in a temporal window preceding that spike, N is the total number of spikes in the analysis. Based on visual inspection of the RevCors, we set

the analysis window to be 15 ms, i.e., we consider that impulse responses are shorter than 15 ms.

The neurons RevCors are first fitted using gamma tone functions (Patterson, 1994). The gamma tone is a cosine carrier with a gamma envelope:

$$GT(t) = A\,(t-t_0)^3 \exp\left(-\frac{t-t_0}{\tau}\right)\cos\left(2\pi f_0 (t-t_0)+\theta\right)H(t-t_0)$$

where A is a scaling factor, t_0 is a pure delay, τ defines the temporal width of the gamma envelope, f_0 is the center frequency of the carrier, θ is a phase shift, and $H(t)$ is the Heaviside function.

Beside the simple gamma tone function, we also fit functions of which the carrier is a chirp, i.e., a frequency modulated signal (Wagner et al., 2009; Fischer et al., 2011). We consider two types of chirps. The first one is based on measurements reported in the auditory nerve of cats. Its instantaneous frequency increases linearly with time (Carney et al., 1999), we refer to it as the linear gammachirp GC_{lin}:

$$GC_{lin}(t) = A\,(t-t_0)^3 \exp\left(-\frac{t-t_0}{\tau}\right)\cos(2\pi\left[f_0 (t-t_0)+0.5c(t-t_0)^2\right]+\theta)H(t-t_0)$$

In the second one, proposed by (Irino and Patterson, 2001), the instantaneous frequency saturates when *t* grows to infinity. We refer to it as the logarithmic gammachirp GC_{log}:

$$GC_{log}(t) = A\,(t-t_0)^3 \exp\left(-\frac{t-t_0}{\tau}\right)\cos(2\pi\left[f_0 (t-t_0)+c\,log(t-t_0)\right]+\theta)H(t-t_0)$$

In both GC_{lin} and GC_{log} the additional parameter c characterizes the rate of the chirp. As the instantaneous frequency f_{inst} is defined as the temporal derivative of the phase, $f_{inst} = f_0 + ct$ for GC_{lin} and $f_{inst} = f_0 + c/t$ for GC_{log}. Thus, f_0 can be seen as the starting frequency of the chirp in the linear case, and as the frequency to which the carrier converges as t grows in the logarithmic case.

These functions are fitted to the RevCors in order to minimize the error $\chi^2 = \frac{1}{N-M}\sum_{i=1}^{N}\frac{(h(t_i)-g(t_i))^2}{\sigma_i^2}$, where $g(t)$ refers to either $GT(t)$ or $GC(t)$, N is the number of time points, M is the number of parameters to fit, and σ_i^2 is an estimate of the variance of the RevCor at time point t_i across all presented trials.

We quantify the effect of stimulus level on the resulting parameters by computing the percentage of change per dB, for each neuron with more than one stimulus level recorded. The fitting procedure yields a set of fitted parameters at each level. For each parameter, we perform a linear regression between the stimulus level in dB and the fitted parameter value. The slope of this regression is the level sensitivity of the corresponding parameter. This slope is divided by

197 the mean parameter value across levels, and multiplied by 100 to yield the percentage of change
198 per dB.

199 Similarly to models of the auditory periphery based on RevCor filters (de Boer and de Jongh,
200 1978; Patterson, 1994), the input sound stimulus $s(t)$ is processed by a FIR filter with an
201 impulse response k that is the truncated version (of length 15ms) of the fitted function:
202 $x(t) = k * s(t)$, where * denotes the convolution operator. The signal is then delayed by a certain
203 amount of time Δ, to compensate for delays introduced by subsequent stages of the model.

204

205 ***Spiking neuron models***

206 The first phenomenological spiking neuron considered (Fig. 5B) is the leaky integrate-and-fire
207 (LIF) neuron, which has been shown to efficiently model responses of a wide class of neurons
208 (Jolivet et al., 2004; Gerstner and Naud, 2009). The output *x* of the auditory filter is first half-
209 wave rectified and compressed by a power law: $I(t) = [x(t-\Delta)^+]^c$ with c chosen between 0
210 and 1. The resulting signal I(t) is then fed to the LIF. The subthreshold membrane voltage
211 dynamic of a LIF neuron is described by a first-order linear differential equation:

212 $$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + I(t)$$

213 where $V_m(t)$ is the membrane voltage, τ_m is the membrane time constant, and $I(t)$ is the input
214 current. The neuron fires when $V_m(t)$ exceeds a fixed threshold V_T. After firing, the membrane
215 voltage is reset to 0: $V_m \to 0$ and the neuron cannot fire during a fixed refractory period *r*. The
216 second spiking model considered is a variation of the spiking model for stimulus level-invariant
217 processing recently proposed in (Brette, 2012) (inset Fig. 6E). We call this model the adaptive
218 threshold model (ATM). As before, the output x of the filter is half-wave rectified but *not*
219 compressed: $I(t) = x(t-\Delta)^+$. Next, $I(t)$ is directly compared to a threshold $V_T(t)$ which can
220 vary in time. The dynamics of the threshold is described by a first order differential equation,
221 which linearly depends on $I(t)$:

222 $$\tau_T \frac{dV_T(t)}{dt} = aI(t) - V_T(t)$$

223 where V_T is the time varying threshold, τ_T is the threshold time constant, and a quantifies the
224 amount of subthreshold adaptation of V_T. A spike is fired if the input exceeds the threshold:
225 $I(t) > V_T(t)$. After firing, the threshold is reset: $V_T \to \beta V_T + \alpha$ and the neuron cannot fire
226 during a fixed refractory period *r*. This reset consists of two parts, an additive part α and a
227 multiplicative part β. In (Brette, 2012), it was proven that a purely multiplicative reset ($\alpha = 0$)
228 yields a level-invariant neuron model, i.e., after a transient time, spike timing and firing rate do
229 not change with stimulus level. In order to account for some level sensitivity seen in the
230 recorded responses, we include an additive term α. In some cells, adding a second time
231 constant to the process yields better results. To do so, a second threshold equation with the

same reset mechanism but different parameter values is introduced, and the condition for firing is then $I(t) > V_{T1}(t) + V_{T2}(t)$. The inclusion of a second time constant does not change the theoretical properties of the model (Brette, 2012). Note that, contrary to the LIF, the ATM directly compares the input $I(t)$ with the threshold. We made this choice because using $V(t)$ as input to the threshold equation would not yield different results (Brette 2012), while adding an extra parameter to the model.

Having an adaptive threshold is not a new concept (Brandman and Nelson, 2002; Chacron et al., 2003; Kobayashi et al., 2009). Nevertheless, in those previous models, the threshold did not adapt to the subthreshold potential and the reset was purely additive. This model can be seen as a special case of the ATM with $a = 0$, $\beta = 0$, and $\alpha \neq 0$. For the sake of comparison we also test this simple adaptive model.

Model fitting procedure

There are several parameters to find in order to optimize the two models: τ_m, c, V_T, r, and Δ for the LIF model, and Δ, τ_T, a, α, β, r for the adaptive threshold model. The model fitting approach employed for the optimization is similar to the one introduced in (Rossant et al., 2010, 2011). In order to quantify the similarity between two single spike trains, we first use a measure that takes into account the precise timing of spikes given a temporal window δ, the gamma factor Γ (Jolivet et al., 2008):

$$\Gamma = \left(\frac{2}{1 - 2\delta r_{exp}} \right) \frac{N_{coinc} - 2N_{exp}\delta r_{exp}}{N_{exp} + N_{model}}$$

where r_{exp} is the mean firing rate of the experimental response, N_{coinc} is the number of coincidences between the model and experimental trains computed within a time window δ, N_{exp} and N_{model} denote the number of spikes in the experiment and model spike train, respectively. $2N_{exp}\delta r_{exp}$ is the expected number of coincidences generated by a Poisson process with rate r_{exp}. The first term in brackets is a normalization factor so that the maximum of Γ is 1. $\Gamma = 0$ means that there are no more coincidences than expected by chance whereas $\Gamma = 1$ means that the model prediction is perfect, at temporal resolution δ. For each cell, there exists a maximum for Γ at a given δ, $\delta_{\max}$ (0.5±0.16 ms for the entire population). For the optimization fitness, we set $\delta = \delta_{\max}$ for each cell . All the optimization results are consistent as long as the chosen δ remains in the vicinity of $\delta_{\max}$, which can be seen as the optimal temporal resolution to compute Γ. If they differ too much from $\delta_{\max}$ (more than a millisecond), the optimizations fail.

The model only outputs one spike train per frozen noise, since it is deterministic, whereas the data contains several repetitions of the same frozen noise. Therefore, we calculate the gamma factor $\Gamma(model, data)$ between the model and the data as the mean Γ between the model

train and each train of the data, i.e., $\left\langle \Gamma\left(model, data\right)\right\rangle = \frac{1}{n}\sum_{k=1}^{n}\Gamma_k$, where Γ_k is the gamma factor between the model train and the k th train out of n trials. Another useful metric that we will use is the intrinsic gamma factor of a set of repeated trials, $\Gamma_{int}\left(data\right) = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\Gamma_{ij}$, where Γ_{ij} is the gamma factor between the trains i and j out of the n trials. It quantifies the reproducibility of responses.

We will use a fitness criterion that takes into account both the quality of spike timing prediction and of firing rate prediction:

$$fitness\left(model, data\right) = \frac{\left|\Gamma\left(model, data\right) - \Gamma_{int}\left(data\right)\right|}{\Gamma_{int}\left(data\right)} + \lambda\frac{\left|FR\left(model\right) - FR\left(data\right)\right|}{FR\left(data\right)}$$

In theory the difference in firing rates is taken into account in the gamma factor. For some cells the regularization factor helped the optimization algorithm to quickly find a relevant parameter subspace. A regularization weight λ of 0.2 was empirically found to give fast convergence. The final results were not sensitive to this value. The optimization uses an evolution algorithm called CMAES (Hansen and Ostermeier, 2001). The implementation on Graphical Processing Units (GPU) is described in (Rossant et al., 2010, 2011). All the neuron simulations and optimizations were performed using the Brian simulator (Goodman and Brette, 2009) for spiking neuron models, the Brian Hears toolbox for auditory filtering (Fontaine et al., 2011) and the Playdoh optimization toolbox (Rossant et al., 2011). All simulations were performed with a sampling frequency of 65 kHz.

Training and testing were done on distinct subsets of the data. When only one 1-second stimulus noise was available, the first 500 ms were used for training, whereas the last 500 ms for testing. To discard the transients when testing, the simulation started at 400 ms, but the testing performances were computed from 500 ms on. When two 1-second stimuli were available, the first stimulus was used for learning and the other one for testing. To compute the fitness, the first 50ms were discarded. Two learning protocols were used: equal level learning and multiple levels learning. In equal level learning, a model is optimized for each level of a cell, yielding as many fitted models as there are levels. The testing is then done at each level using the model learned for this level. In multiple levels learning, only one model is learned for each cell. All the responses from the learning data set are concatenated with 100 ms silence between successive responses, and the fitness is computed over the whole response. The testing is then performed on every single level of the cell using the same learned model. To compute correlations in the recorded dataset, spikes from different trials are used, i.e. 50 trials of a 1-second stimulus. Because our models (LIF or ATM) yield identical trains for different trials we used a 50-second fresh noise to obtain a sufficient number of spikes and computed the cross-correlation on the modelled data at different levels.

Different metrics were used to compare model and recorded responses. In particular, differences between firing rates, HHWs, CIs, response lags, as well as between $\Gamma\left(model, data\right)$ and $\Gamma_{int}\left(data\right)$, were quantified using an explained variance (EV) measure defined as:

306 $EV = 1 - SS_{err} / SS_{tot}$ where the square of the error $SS_{err} = \sum_i (y_i - \hat{y}_i)^2$ and the empirical

307 variance $SS_{tot} = \sum_i (y_i - \bar{y})^2$ with $\bar{y} = 1/n \sum_i^n y_i$. Here, y_i is the associated metric to the recorded

308 response *i* and $\hat{y}_i$ the metric associated to the corresponding modeled response. Note that

309 when the fit is really bad the variance of the error can be larger than the empirical variance,

310 resulting in a negative explained variance.

311

312 ***Stochastic adaptive threshold model***

313 In order to account for stochasticity in the data set, noise is added to the model. This is simply

314 done by adding a white noise term to the threshold equation. The standard deviation of the

315 noise scales linearly with stimulus level, so that the signal-to-noise ratio is constant.

316 $$\tau_T \frac{dV_T(t)}{dt} = aI(t) - V_T(t) + \bar{I}(t)\sigma_T \sqrt{2\tau_T}\xi(t)$$

317 where $\xi(t)$ is gaussian noise, σ_T is the level-independent standard deviation of the noise, and

318 $\bar{I}(t)$ is a running average of the input: $\tau_T \frac{d\bar{I}(t)}{dt} = -\bar{I}(t) + I(t)$. We set τ_T to 20ms; σ_T is

319 optimized so that the main lobe of the model response SACs at different levels match the main

320 lobe of the SACs of the corresponding recorded data at the same levels (using a mean square

321 error criterion). When the threshold has two dynamic equations (two time constants), the same

322 noise is added to both of them.

323

324 ***Linear non-linear Poisson model***

325 We compared our stochastic model with a popular model, the linear-nonlinear-Poisson (LNP)

326 model (Chichilnisky, 2001; Pillow et al., 2005). Similar to the two models previously introduced,

327 the input to our LNP model is the filtered sound stimulus. This input is passed through an

328 instantaneous nonlinear function f, which accounts for non-linearities such as rectification and

329 saturation. The instantaneous spiking probability in response to a stimulus s is as follows:

330 $$P(spike(t)|s) = f(k * s(t - \Delta))$$

331 where $s(t)$ is the stimulus, Δ a time delay, k the auditory filter impulse response estimated by

332 reverse correlation, * denotes the convolution operator, and f is a non-linear function to

333 optimize. Spikes are produced using an inhomogeneous Poisson process. To avoid non-realistic

334 bursting, a refractory period was set to 1 ms which is smaller than the shortest characteristic

335 period (1/CF) considered.

336 Using Bayes' rule, the non-linearity function $f(s) = P(spike|s)$ can be rewritten as

337 $f(s) = \alpha P(s|spike)/P(s)$. The prior P(s) can be estimated using a Gaussian kernel density

estimate from the stimulus ensemble. Similarly, $P(s \mid spike)$ was estimated from the spike-triggered stimulus ensemble, i.e., the stimulus values at spike times. α was optimized so that the firing rate of the model fit the firing rate of the cell.

Predicting spike count reliability

We want to relate the spike count reliability of a response to the distance between the input stimulus and the threshold of a model neuron. To do so, we use a standard definition of the reliability of a response during a stimulus event. First, events are defined as time intervals where the input to the cell is positive, i.e. each positive "chunk" of the filtered sound. Because the cell's input has a characteristic period induced by the filtering, events are well separated in time (see Fig. 12 A). When a frozen noise is presented *n* times to a neuron, the reliability for event *i* is defined as the number *p* of trials in which the cell spiked during this event divided by the total number of trials, $R_i = p/n$ (Mainen and Sejnowski, 1995). An event reliability of zero means that no spike has been fired whereas a value of one means that a spike was fired in every trial.

The distance between the input *I(t)* and the dynamic threshold at the *i*th event is given by the difference between the peak magnitude of *I(t)* in the time interval defining the event *i* and the value of the average dynamic threshold over all trials <V(t)> at the beginning of the event *i* (e.g. in Fig. 12A for the stochastic ATM), The distances and peak magnitudes are normalized with respect to the mean stimulus level. Distance, peak magnitude, and reliability *R,* are computed for every event, for a given neuron and input level. The resulting pairs (x_i, y_i) (reliability-distance or reliability-peak magnitude) are fitted, using a least-square method, to the sigmoidal function $f(x) = 0.5\left(1+\mathrm{erf}\left(\frac{x-u}{\sqrt{2\sigma}}\right)\right)$, where erf is the error function. This is the cumulative distribution function of a normal distribution with mean u and variance σ^2. For each reliability value x_i, this procedure yields an estimation $f(x_i)$ of the distance or the peak magnitude (see e.g. Fig. 12B). To quantify the quality of the fit, the coefficient of determination is computed as $R^2 = 1 - SS_{err}/SS_{tot}$, where the fit squared error $SS_{err} = \sum_i (y_i - f(x_i))^2$ and the total empirical variance $SS_{tot} = \sum_i (y_i - \bar{y})^2$ with $\bar{y} = 1/n \sum_{i=1}^{n} y_i$.

Spike effect on spiking probability

If an adaptive threshold is involved in the spike generation process, the firing probability of spikes at time *t* should depend on the occurrence of preceding spikes at times $t_0 < t$. To test whether such an effect is present in the TB responses, we calculate, for every stimulus event *i* generating at least one spike, probabilities of firing depending on spike history. The procedure is illustrated in Fig. 7A where events are the intervals between dashed lines. We first calculate the

probability that a spike is generated at time t (t in event *i*) given that a spike occurred in a given past temporal window Δ (green in Fig. 7A):

$$P^{i}_{1\to1} = P\left(spike\,at\,t \in \text{ event } i | spike\,in\,\Delta\right) = \frac{N_{11}}{N_{10}+N_{11}},$$

where N_{11} is the number of trials in which there was a spike both in the preceding window $\Delta = [t-t_a, t-t_b]$ *and* in event *i* (*t* represents the beginning of the event , green box in Fig. 7 A. N_{10} is the number of trials in which there was a spike in the preceding window Δ *but not* in event i. Similarly, we define the probability for the *i*th event that a spike is generated given that no spike previously occurred in Δ :

$$P^{i}_{0\to1} = P\left(spike\,at\,t \in \text{ event } i | no\,spike\,in\,\Delta\right) = \frac{N_{01}}{N_{00}+N_{01}}.$$

N_{01} is the number of trials in which no spike occurred in Δ but a spike occurred in event *i*, and N_{00} is the number of trials in which there was neither a spike in Δ nor in event *i*. For a given event, if spikes that occurred in the past in Δ have a suppressive effect on subsequent spikes, then $P^{i}_{0\to1} > P^{i}_{1\to1}$. In order to discard possible effects of the refractory period, for a given event starting at time *t*, spikes occurring between *t* and *t-r* , where *r* is the refractory period, are discarded (gray box in Fig. 7A). The refractory period is defined as the shortest interval where the inter-spike interval histogram exceeds 5% of its maximum (Fig. 7B). This procedure is repeated for every event of every response. For visualization (Fig. 7C and D), all points ($P^{i}_{0\to1}, P^{i}_{1\to1}$) are used to estimate the joint probability density via a 2-dimensional kernel density estimator using a Gaussian kernel. To ease visualization, each column is normalized to its maximum.

Results

In this study we model low-frequency cochlear nucleus neurons that are highly synchronous (high-sync), i.e., cells that generate spikes that are precisely timed to the fine structure of sounds (Joris and Smith, 2008). From visual inspection of the raster plots (e.g. in Fig. 1), we define the following selection criteria: we select low-frequency cells (CF<1000Hz) with at least one level for which the responses are reliable enough (CI>5). All these cells were classified as PHL. Based on the PSTH and recording location the recordings were likely from axons of bushy cells (Joris et al., 1994). The final data set used in this study contained 24 cells, 4 of which were recorded at a single SPL, 8 of them at 2 SPLs, and 10 at more than 3 SPLs. Stimulus levels were separated by at least 20 dB. For all cells at least one 1-second frozen noise was presented between 30 and 50 times. When time allowed (n=10), another 1-second frozen noise (a different token than the first one) was presented, with the same number of repetitions

High-sync response properties across stimulus levels

From the autocorrelation analysis (Fig. 2A, see also Material and Methods), we estimate the temporal precision (Fig. 3A) and the correlation index (CI) (Fig. 3B) of responses at each level. Connected points correspond to the same cell at different levels. First, we note that the responses are very precise: half-height width (HHW) is smaller than 1 ms (population median is plotted in red as a function of level in Fig. 3A) for almost all cells and all levels. Second, as was shown in (Louage et al., 2005), individual HHWs and correlation indexes tend to decrease with increasing level . Fig. 3C shows the firing rate as a function of stimulus level for all cells and the population median (in red). The levels range from 40 dB to 110 dB, which corresponds to a change in input pressure by a factor of about 3000. Thus the firing rate responses exhibit strong compression.

Another standard measure of the temporal precision of the responses is the intrinsic gamma factor Γ_{int} (Jolivet et al., 2008), which quantifies the coincidences between responses across trials, at a given temporal resolution δ. $\Gamma_{int} = 0$ means that there are no more coincidences between trials than expected by chance for a Poisson process and $\Gamma_{int} = 1$ means that all the trials are identical at the given time resolution. Fig. 3D shows Γ_{int} for all cells and levels, computed at resolution δ = 0.5 ms and the population median (in red). This figure mirrors the trend seen in Fig. 3A, that is, the temporal precision of responses is enhanced at higher levels.

A standard measure of spike count reproducibility is the Fano factor (FF), defined as the variance of spike count divided by its mean. It equals 1 for a Poisson process. We computed the average FF over a sliding window of 30 ms, for every cell and level in our dataset (Fig. 3E). The population median as a function of level is shown in red. Similarly to what has been reported in various sensory systems (auditory system (Young and Barta, 1986; Avissar et al., 2007), and other systems (Berry et al., 1997; de Ruyter van Steveninck et al., 1997)), the TB responses exhibit sub-Poisson properties (FF<1), that is, responses are more reproducible than for Poisson processes. The FF decreases as the stimulus level increases.

As the high-sync cells - presumably bushy cells - convey temporal information to binaural cells sensitive to ITD, it seems functionally useful that their responses be as insensitive to stimulus level as possible. In other pathways, an often observed effect of stimulus level on spike trains is a temporal shift, with shorter spike latency at higher level (Gollisch and Meister, 2008). If this effect occurred for the monaural inputs of the binaural cells, then the ITD tuning of binaural cells would depend on interaural level differences (ILD) (Brette, 2012). We measured this temporal shift as a function of stimulus level by calculating the XAC between the responses at a given SPL (Fig. 2B, see Material and Methods) and the responses at a reference level (chosen as 70 dB when available, 60 dB or 80 dB otherwise). The temporal shift of the cross-correlogram peak characterizes the lag between the corresponding responses (Fig. 2B3). The results are shown in Fig. 3F where it can be seen that, except for one fiber, the variation of the response lags hardly exceeds 200 µs, while stimulus level varies by more than 40 dB. In Fig. 3G, the same results are shown in cycles, i.e., temporal lag multiplied by the CF of the cells. The sensitivity of each cell, defined as the slope of the linear regression of the lag/SPL relationship, is plotted in Fig. 3H in µs/10dB and in Fig. 3I in cycle/10dB. It appears that spike timing in these cells is not very sensitive to input stimulus level (median = 16µs/10dB in Fig. 3H, = 0.9×10^{-2}cycle/10dB in Fig.

3I) when considering the natural range of ILDs. Indeed, at these frequencies, the maximum ILD is about 5 dB (Tollin and Koka, 2009), yielding a median lag of 8 µs (see Discussion).

Reverse correlation analysis

We estimated the reverse correlation filters of the neurons from responses to broadband noises using a reverse correlation (RevCor) technique (see Materials and Methods). Examples of RevCors at 6 different SPLs for the same neuron are shown in Fig. 4A. Note that the stimuli used for the analysis were normalized to have unit power so that the amplitudes of the resulting RevCors have the same order of magnitude. The RevCors do not vary much with stimulus level. This can be quantified for each neuron by calculating the maximum of the cross-correlation function between every pair of RevCors at different levels, and by averaging across all possible level pairs. On average, the RevCors of a cell at different levels are highly correlated with each other (0.91±0.02), indicating that the RevCor shape does not vary much with input level in most of the cells.

We fitted the RevCors to functions with gamma envelope and different carriers, chirping and non-chirping (see Materials and Methods). The RevCors were better fit by linear or logarithmic gammachirps than by gammatones (Mann-Whitney U test, $P<0.02$ in both cases). This finding is consistent with previous studies in the cochlear nucleus of barn owls (Wagner et al., 2009; Fischer et al., 2011) or in the auditory nerve of cats (Carney et al., 1999). We found no significant difference in estimation error between the logarithmic chirp function and the linear one (median linear chirp χ^2 =0.043, median logarithmic chirp χ^2 =0.044, P=0.42, Mann-Whitney U test) . Fig. 4B shows the linear gamma chirp functions fitted to the RevCors of Fig. 4A.

To further quantify the effect of stimulus level on the RevCor shapes, we analyzed the parameters of the fits (Fig. 4 C,D, and E). The left column shows the envelope widths τ (C), starting frequencies of the chirps f_0 (D), , and gliding slopes c (F), for every cell and every level as a function of their CF. Each cell is represented by a shape (circle, square, pentagon, and triangles with different orientation) and connected points correspond to the same cell at different levels. Level is color-coded for each neuron with the darkest color for the lowest level and the brightest color for the highest level. The envelope width τ is inversely correlated with CF (mean: 0.88 ms, regression -10^{-4} ms/Hz x+1.36ms, r = -0.55, $p<10^{-5}$). The starting frequency of the chirp is positively correlated with CF (mean: 633 Hz, regression: 0.8 x+261 Hz, r = 0.66, $p<10^{-8}$). The glide slope c is not correlated with CF (mean: -0.05 kHz/ms, regression 3×10^{-5} kHz^2/ms x -0.06 kHz/ms, r = 0.08, p<0.53), neither is the phase (mean: 0.62 cycle, regression: 8×10^{-5} cyc/Hz x+0.58 cycle, r = 0.06, p = 0.61)

In order to quantify the effect of stimulus level on parameter values, we plot the percentage of change per dB, relative to the average value of the parameter (Fig. 4 C,D, and E, right column, see Materials and Methods). For most of the cells, the effect of stimulus level is small. The level sensitivity of τ is small and negative (mean % of change/10dB = -02.5+-5 %), the sensitivity of f_0 is also small but positive (mean % of change/10dB = 5+-5%). The sensitivity of the gliding slope is higher due to one outlier (mean % of change/10dB = 26+-72%) but overall is also mainly level-independent, as reported previously for AN fibers (Carney et al., 1999) and basilar

membrane (de Boer and Nuttall, 1997), but see (Recio-Spinoso et al., 2009). The level sensitivity of the phase θ is small and around zero (mean % of change/10dB = 3+-19%).

From these observations, we can conclude that the stimulus level has little effect on the shape of the RevCors. For the rest of this study, the filter used in the model of each neuron (Fig. 5A) is the linear gammachirp function fitted to the RevCor obtained at a given reference level. This reference level is 70 dB when available (n=16), 60 dB (n=1) or 80 dB (n=1) otherwise.

Equal level learning with a simple integrate-and-fire model

We first study the predictive power of a simple spiking neuron model, a LIF model with fixed threshold and compression (Fig. 5 A and B, see Materials and Methods). Learning and testing are done at the same level, but with different stimuli. That is, for each cell, there are as many fitted models as levels. Since this simple LIF model is deterministic, it produces the exact same spike trains in response to a given stimulus in all trials, while the responses are variable in the data. The model is optimized so that 1) the spike train produced by the model is maximally coincident with the spike trains in the data, at a resolution of 0.5 ms, and 2) the firing rate of the model is similar to the average firing rate of the data (Fig. 5 A). Fig. 5C shows the responses of a cell at two different levels (dots) and the two spike trains produced by the fitted model (red), on a test stimulus - i.e., a different stimulus was used to fit the model. The model appears to predict spike times with good accuracy in this example. We note that the model misses a few volleys of spikes, especially at the lower level. There is about one volley of spikes for each characteristic period, but the firing rate of the cell is lower than the characteristic frequency (CF=462Hz, firing rate =180Hz for 50dB and 240Hz for 70dB). Thus, on any given trial, the cell does not fire on each period of the stimulus, and the same is true for the model.

We now examine the prediction performance on the whole population, for every neuron and every level (Fig. 5D, E). Each situation corresponds to a specific set of parameter values. Fig. 5D shows a very good agreement between the model firing rate and the firing rate of the data (correlation coefficient $R = 0.91$, explained variance $EV = 0.62$). Fig. 5E shows that spike timing is also well predicted: the gamma factor between model spike trains and data spike trains $\Gamma(data, model)$ is close to the intrinsic gamma factor of the data $\Gamma_{int}(data)$ ($R = 0.88$, $EV = 0.64$). The statistics of the resulting parameters are shown in Table 1 for all cells at all levels.

Parameter	τ_m	c	V_T	Δ	r
Initial range	(0.05ms, 20ms)	(0, 1)	(0.01, 15)	(-2ms, 2ms)	(0.1ms, 10ms)
Result, mean±SD	1.2±0.8 ms	0.55±0.16	0.95±0.92	-0.05±0.26ms	2.5±1.6ms

Table 1: Fitted parameter values for the LIF model learned at single stimulus levels (n=24).

As we previously noted, there is one set of parameter values for each level. In particular, the fitted values depend on level. Fig. 5F and G show how the spike threshold V_T and the compression exponent depend on level. We note that the spike threshold increases steeply with stimulus level.

Predicting responses across levels

We now consider the more realistic case where there is a single set of parameter values for each cell, regardless of the stimulus level. That is, the model must predict the cell's responses to all stimuli, with no a priori knowledge of the stimulus level. The learning set consists of concatenated responses at the available stimulus levels, separated by silent periods of 100 ms (see Materials and Methods). The model is fitted on this learning set and tested at each level, with different stimuli.

In this new condition, the fixed threshold model performs poorly (Fig. 6). An example is shown in Fig. 6A, with the responses of the fitted model at 6 different levels (green) superimposed on the cell's responses. In this example, the model does not fire at the lower levels (40 dB and 50 dB). It can be seen in Fig. 6B that the model (green) fires more than the cell (blue) at higher levels and less than the cell or not at all at lower levels. A second observation is that the model tends to fire too early at higher levels and too late at lower levels. This is shown quantitatively in Fig. 6C, where the lag of the responses with respect to a reference stimulus level of 70 dB is shown for the model and for the cell.

The reason for this poor performance is suggested in Fig. 5F: to correctly predict responses across levels, the spike threshold must increase with level. Fig. 6D illustrates what happens when the spike threshold is fixed. When the stimulus level increases, the threshold (red) is crossed earlier and therefore spikes are produced earlier. In addition, previously sub-threshold events may become supra-threshold and new spikes may appear. Conversely, when the stimulus level decreases, spikes are produced later and a few may disappear.

From these considerations, we conclude that the threshold should adapt to the input in order to reduce the effect of level. Our starting point is the adaptive threshold model (ATM, inset Fig. 6E) recently proposed in (Brette, 2012) that has level-invariant responses, both in terms of spike timing and firing rate. This model is based on the observation that, to produce level-invariant responses, scaling the input should leave the crossing points unchanged (Fig. 6E). This constraint implies that the threshold should depend linearly on the input, and the increase in threshold following a spike must be multiplicative, i.e., it must be proportional to the threshold value at spike time. In order to take into account deviation from complete level invariance, an additive term is added to the reset (see Material and methods).

When the same single set parameter optimization procedure is applied to this adaptive model, the prediction performance across levels drastically improves (red line Fig. 6A and F). Both the firing rate (Fig. 6B, red) and spike timing (Fig. 6C, red) are accurately predicted across level. This model has only one more parameter than the LIF model (6 parameters vs. 5), and therefore this drastic increase in performance is not simply the result of an increased complexity.

If a spike-dependent adaptive process were at play, as opposed to e.g. mechanical compression in the cochlea or synaptic depression, the firing probability of spikes at time *t* should depend on

the occurrence of preceding spikes at times $t_0 < t$. Let us define *events* as time intervals where the input to the model cell is positive. To test whether such an effect is present in the TB responses, we calculate, for every stimulus event *i* generating at least one spike, firing probabilities depending on spike history (see Materials and Methods). First, we calculate the probability $P^i_{1\to1}$ that a spike is generated at time t (t in event *i*) given that a spike occurred in a given past temporal window Δ (green box in Fig. 7A). Second, we compute the probability $P^i_{0\to1}$ for the ith event that a spike is generated given that no spike previously occurred in Δ. For a given event i, if spikes that occurred in the past in Δ have a suppressive effect on subsequent spikes, then $P^i_{0\to1} > P^i_{1\to1}$. In order to discard possible effects of the absolute refractory period, for a given event starting at time t, spikes which occurred between t and t-R, where R is the refractory period, are discarded (gray box in Fig. 7A). The absolute refractory period is defined as the shortest interval where the inter-spike interval histogram exceeds 5% of its maximum (Fig. 7B) yielding 1.4±0.4ms for the population. This procedure is repeated for every event of every response, and we plot the 2-dimensional density $P^i_{1\to1}$, $P^i_{0\to1}$ for two different time windows Δ. The first one is $\Delta = [t, t-CP]$, where CP is the characteristic period CP of the neuron and t is the starting point of the event (Fig. 7C). The second one is one period earlier: $\Delta = [t-CP, t-2CP]$ (Fig. 7D). We can see in Fig. 7C that $P^i_{1\to1}$ is significantly lower than $P^i_{0\to1}$ (most of $P^i_{1\to1}$ falls under the diagonal), which is not the case in Fig. 7D. This shows that spikes have a suppressive effect on subsequent spikes for a time of about one characteristic period, and that this effect is not due to the refractory period.

Population analysis of multiple level models

Fig. 8 shows the testing performance of three different models on the entire population, when there is a single set of parameter values for all stimulus levels for each cell, i.e., the learning set consists of concatenated responses at all stimulus levels. Models are tested at all levels on all cells. The first row shows the performance of the fixed threshold LIF model. As expected from the aforementioned considerations, the model tends to have a higher firing rate than the cells at high levels (Fig. 8A1, bright colors) and lower firing rate at low levels (dark colors), yielding poor prediction performance ($R = 0.87$, $EV = -1.51$). In fact, the model responds only for 61% of the stimulus conditions. As a consequence, the similarity between modeled and recorded spike trains is low on average (Fig. 8B1, $R = 0.72$, $EV = 0.1$). The lag of the responses with respect to a reference level (generally 70dB) is plotted in Fig. 8C1. The results on the entire population follow the trend shown in Fig. 8: the responses of fixed threshold models tend to lead the recorded responses at high level and to lag behind them at low level, yielding poor prediction performance ($R = 0.64$, $EV = -8.25$).

In the second row, we show the results for a simpler adaptive threshold model (Brandman and Nelson, 2002; Chacron et al., 2003; Kobayashi et al., 2009), where $a = 0$, $\beta = 0$, and $\alpha \neq 0$. Note that we do not include compression. That is, adaptation is only triggered by spikes, and the threshold increases by a fixed quantity after a spike. We refer to it as simpler ATM (sATM). Although it performs better than the fixed threshold model, it still shows the same problems: the

firing rate is too high at high level and too low at low level (Fig. 8A2, $R = 0.84$, $EV = 0.06$), and spike timing prediction is poor (Fig. 8B2, $R = 0.82$, $EV = 0.28$), because it tends to fire too early at high level and too late at low level (Fig. 8C2, $R = 0.61$, $EV = -1.92$).

In the third row, we show the results of the purely level invariant model (Brette, 2012) where a≠ 0 , β ≠ 0 , and α = 0. As expected, the responses are the same across levels. The firing rates do not vary with intensity (the points on Fig.8A3 are placed on a horizontal line for each cell). Because spike timing is also level invariant, the lag is zero for all levels compared to the reference level (Fig.8C3). Even if this model is capable of predicting the response to some degree (Fig.8B3), those results confirm the fact that a level-dependent term must be included (α ≠ 0). This is done in our ATM model, for which the results (fourth row) show a very good match between the firing rate of the model and the firing rate of the recorded responses (Fig. 8A4, $R = 0.97$, $EV = 0.92$). The model also shows very good performance in predicting the spike trains (Fig. 8B4, $R = 0.95$, $EV = 0.68$). Fig. 8C4 shows that the lag of the responses as a function of level is very similar between the model and the data ($R = 0.82$, $EV = 0.58$), with a regression line close to the diagonal. These results show that the ATM is better at predicting responses across level than a model with fixed threshold, even though the fixed threshold model included compression. In addition, our ATM model, which includes subthreshold adaptation and multiplicative reset, significantly improves upon a sATM.

Statistics of the optimized parameters for the ATM are given in Table 2. A few cells (n=5) were better fit with a threshold consisting of two dynamical processes with two different time constants. Statistics for these cells are shown in Table 3.

Parameter	a	α	β	Δ	τ_T	r
Initial range	(0, 20)	(0, 10)	(0.5, 20)	(-2ms, 2ms)	(0.5ms, 80ms)	(0.1 ms, 10 ms)
Mean+-SD	3.2±1	0.8±0.9	4.1±3	0.02±0.005 ms	5.3±4.9 ms	2.2±0.8 ms

Table2: Fitted parameter values for the ATM learned at multiple stimulus levels (n=15)..

Parameter	$a1$	$\alpha 1$	β1	τ_{T1}	$a2$	$\alpha 2$	β2	τ_{T2}	Δ	r
Mean+-SD	0.5±0.2	2.4±2.4	12.3±5.8	0.7±0.03 ms	40±4ms	0.6±1.2	1.6±1.2	40±34ms	0.01±0.004ms	2±0.4ms

Table3: Fitted parameter values for the ATM with two time constants learned at multiple stimulus levels. The first set of parameters corresponds to the faster time constant, the second one to the slower time constant. The initial ranges were the same as in Table 2 (n=5).

Stochastic models

So far, we only considered deterministic spiking neuron models, i.e., with no intrinsic noise. Although the *in vivo* responses of TB fibers are temporally precise and reliable, there is still some variability, both in timing and in spike count. To account for this variability, we now add a white Gaussian noise, with a given variance σ_{noise}^2, to the spike threshold (see Materials and Methods). To maintain a constant signal-to-noise ratio, the standard deviation σ_{noise} is proportional to the stimulus level: $\sigma_{noise} = \sigma\bar{I}$, where $\bar{I}$ is a low-pass filtered version of the input I. The invariant part σ of the noise variance is optimized so that the main lobe of the SACs of the model responses at different levels match the main lobe of the SACs of the corresponding recorded data at the same levels (using a mean square error criterion). We refer to this model as the stochastic ATM.

We first consider the case of single-level learning. We compare our model with a widely used approach in neural modeling: the linear-nonlinear Poisson model (LNP(Chichilnisky, 2001; Pillow et al., 2005)). The LNP model consists of a cascade of a linear and a non-linear stage, followed by Poisson spike generation (see Materials and Methods). The linear part is the same auditory filter as previously used whereas the static non-linearity is optimized on the learning dataset. A 60 ms raster plot of responses from the testing set of a TB fiber at two stimulus levels is shown in Fig. 9, for the recorded TB fiber responses (A), the stochastic ATM (B) and the optimized LNP model (C). In this particular example, the LNP responses show more spike jitter than the data whereas the stochastic ATM responses seem qualitatively more similar.

Prediction performance on the testing set are shown for the entire population in Fig. 10, where the firing rates (Fig. 10A1), HHW (Fig. 10B1), and CI (Fig. 10C1) of the recorded responses (x-axis) are compared with those of the corresponding models (y-axis). The LNP model is better at predicting the firing rate than the stochastic ATM (compare Fig. A1 and A2, $R = 0.99$ and $EV = 0.97$ for the LNP model, $R = 0.97$ and $EV = 0.57$ for the stochastic ATM). However, the precision of spike timing is poorly predicted by the LNP model, which is generally less precise than the recorded cells: the HHW is too high (Fig. B1, $R = 0.61$ and $EV = -1.15$) and the correlation index is too low (Fig. C1, $R = 0.77$ and $EV = 0.07$). On the other hand, the precision of spike timing in the stochastic ATM matches the precision of the data very well (HHW in Fig. 10B2, $R = 0.75$ and $EV = 0.43$; CI in Fig. 10C2, $R = 0.94$ and $EV = 0.8$). We computed the correlation coefficients between the PSTH of the model and the PSTH of recorded responses for the two models. Fig. 10D shows the correlation for the ATM against the correlation for the LNP for all cells and levels. The PSTHs are clearly better predicted by our adaptive model than by the LNP model (2-sided t-test: $p=6\times10^{-8}$, mean correlation coefficients between the data and the model: 0.65±0.15 for the stochastic ATM and 0.49±0.20 for the LNP model). We can conclude that, even in the simple case when learning and testing are performed at the same level, the predictions of the stochastic ATM are better than those of the LNP model, because the LNP model is not temporally precise enough.

As we did for the deterministic case, we now analyze the testing prediction performance of the stochastic ATM when the learning set consists of multiple stimulus levels, that is, there is a single set of parameter values of all tested levels (Fig. 11, n=20). We do not show the results for the LNP model, because they are extremely poor. By construction, the LNP model does not generalize well across levels: the firing rate is very sensitive to level, spike timing is not sensitive

679 at all, and precision decreases (HHW increases) with increasing level. For the stochastic ATM,
680 the prediction performance for the firing rate is shown in Fig. 11A. Although the model slightly
681 overestimates the firing rate, its predictions are good across the entire level range ($R = 0.95$,
682 $EV = 0.82$). The predicted temporal precision is also slightly higher than the precision of the
683 cells (HHWs are lower for the models than for the recorded data) but they are good on average
684 (Fig. 11B, $R = 0.69$, $EV = 0.22$). The CI is also well predicted (Fig. 11C, $R = 0.82$, $EV = 0.73$).
685 Finally, the prediction performance on the response lags is also very high (Fig. 11D, $R = 0.87$,
686 $EV = 0.72$). The linear regression (dashed line) suggests that the model is in general not
687 sensitive enough to stimulus level.

688

689

690 ***Predicting spike reliability***

691 It can be seen in Fig. 9A that for a given stimulus, some spiking events are more reliable than
692 others. By *reliable event*, we mean that spikes are observed in most trials in the corresponding
693 event, which is defined as an interval where the filtered stimulus I(t) is positive (e.g. events E1
694 and E2 in Fig. 12A). For each event, we define reliability R in a similar way as (Mainen and
695 Sejnowski, 1995), as the proportion of trials in which the neuron spikes in the event. The total
696 number of trials ranges between 35 and 100 in our dataset. *R=0* means that no spike was
697 observed in response to the stimulus event (unreliable event), whereas *R=1* means that a spike
698 was observed in every trial, i.e., the response is perfectly reliable.

699 We try to explain the reliability of events using our stochastic adaptive threshold model. In our
700 model, reliability should be higher when the input is near or above the average threshold (event
701 E2 in Fig. 12A) than when it is far below the threshold (event E1 in Fig. 12A): the probability of
702 firing due to the noise is higher in the former than in the latter case. This is indeed seen in the
703 raster plots (Fig. 12A, bottom), both for the model and the data. In a model with a fixed
704 threshold, reliability is expected to be mostly determined by the input amplitude in the event
705 (which correlates with the slope of depolarization). But in this example (Fig. 12A), the input
706 magnitude is higher in E1 than in E2, which suggests that the distance to threshold is a better
707 predictor of reliability.

708 In order to quantify these ideas, we first compute the spike count reliability *R* for each stimulus
709 event, in the responses. We then compute the distance between the mean threshold $\langle V_T(t) \rangle$ and
710 the stimulus *I(t)*, in the corresponding model (d1 and d2 in Fig. 12A). Fig. 12B shows the
711 reliability vs. distance for all events for one cell at a given level. Fig. 12C shows the reliability vs.
712 peak values (maximum of I(t)) for the same cell. Both distance and peak value are normalized to
713 the mean stimulus level. In this example, the distance is a much better predictor of R than the
714 peak values, as indicated by the quality of the fit to a sigmoid function.

715 We then compare the prediction performance of the two quantities, distance and peak, on the
716 entire population (all cells, all levels). For every response and every stimulus level, we calculate
717 the reliability, distance and peak value for all events, and we fit sigmoid functions to the
718 resulting sets of points (one set for reliability vs. distance, one set for reliability vs. peak), as in
719 Fig. 12B, C. This procedure yields two coefficients of determination R^2 for each cell and level, one

for the reliability vs. distance fit and one for the reliability vs. peak fit. We compare these two coefficients across the entire dataset, first for the responses generated by the model (Fig. 12D), i.e., using spikes output by the model and the corresponding threshold., We do the same analysis for the recorded responses (Fig. 12E), i.e., we use recorded spikes and the corresponding modelled threshold. As expected, in the model, reliability is much better explained by the distance to threshold than by the input peak (most points are above the diagonal in Fig. 12D). In the recorded responses, the difference is less clear, but distance is still significantly better at predicting reliability than peak (t-test, $p=0.001$). Given that the threshold was not directly measured but only indirectly inferred through our model fitting procedure, this is an interesting result.

Discussion

In this paper, we presented a phenomenological model of the responses of cochlear nucleus neurons to broadband sounds. It consists of a linear filter followed by an integrate-and-fire model with adaptive threshold. We fitted this model to neuronal data recorded in bushy cell axons of cats, using a recently developed technique (Rossant et al., 2010, 2011). The model predicts the precise timing of spikes produced by these neurons at different sound levels. In particular, it captures an essential property of these neurons: the low sensitivity of spike timing to sound level when considering the natural ILD range. Indeed, when characterizing the detection of ITDs, one must consider the relative timing between both sides. If the absolute timing changes with input level, the relative timing changes by an amount related to the ILD. Acoustical measurements in cat suggest that ILDs at the frequencies studied here are not larger than about 5 dB (Tollin and Koka, 2009). Given that the population median sensitivity to input level is 16 μs/10dB (Fig. 3H) the maximal change in timing across the two ears, expected from ILDs, is about 8 μs (median). This is for positions near the interaural axis where ILDs are maximal. For most spatial positions, the change in timing will be smaller. The lag induced by changes in input level is therefore very small from a behavioral perspective. Note that, due to the paucity of MSO data, it is at present actually unclear whether the level invariance present at the monaural stage of the bushy cells confers invariance in ITD-tuning for ILD and SPL to MSO neurons.

Our approach is similar in aim to previous studies in the visual pathway, e.g. retina (Pillow et al., 2005), where the input-output function is reproduced, but the anatomy is not modeled in details. It does not follow the general trend in modeling of the early auditory pathway. Indeed, in the past two decades great research efforts have led to the development of detailed quantitative models of the auditory nerve response (Zhang et al., 2001; Sumner et al., 2003; Zilany et al., 2009), and of cellular models of neurons of the cochlear nucleus (Kuhlmann et al., 2002; Rothman and Manis, 2003; Zhang and Carney, 2005; McGinley et al., 2012). While those models provide valuable insight of the underlying mechanisms, they were designed to account for a number of properties of auditory nerve fibers or of bushy cells, but not to reproduce the precise spike trains in response to arbitrary input sounds. Making those models predictive would be impractical as they contain several dozens of parameters to be tuned.

We have shown that spiking neuron models with a fixed threshold and a compressive factor or with a simple adaptive threshold with additive reset (Brandman and Nelson, 2002; Chacron et al., 2003; Kobayashi et al., 2009) tended to fire too early at high levels and too late at low levels (Fig. 6 and 8). The motivation for the inclusion of a compression on the input in the LIF was two-fold: 1) to give a fair chance to the IF model, which would be immediately discarded in the absence of any compression, 2) to demonstrate that input compression is insufficient to account for the data. This simple compressive exponent included in the LIF model is not meant to be a realistic model of cochlear compression, which affects both gain and bandwidth (Zhang et al., 2001). In contrast, our model was able to reproduce the effects of stimulus level on neural responses, both for the firing rate and the precise timing of spikes (Fig. 6 and 8). It relies on spike-triggered changes of the threshold that include both an additive and a multiplicative term (Brette, 2012). This model should be useful to build functional models of the auditory system. In particular, spike timing is critical in the ITD processing pathway, and our model provides a simple model of the monaural neurons involved in this neural circuit. This model may also be useful in a neuro-engineering context. Indeed neuromorphic sensors, such as spiking electronic retinas (Lichtsteiner et al., 2006) and cochleas (Liu et al., 2010), also need to address the issue of encoding signals across a large dynamic range.

To account for stochasticity in neural responses, we then added noise in our model, controlled by a single parameter. This noise takes into account the effect of the different sources of noise along the pathway, that is: cochlear, transduction, synaptic, and cellular sources of noise. Similar to previous results (Pillow et al., 2005), we found that a noisy integrate-and-fire model, such as our model, is better at reproducing temporal precision and reliability than an LNP model (Fig. 10). In addition, our stochastic model could predict the neural responses with accurate timing, firing rate, temporal precision, and reliability at different stimulus levels (Fig. 11). Note that the LNP model could be modified to predict responses at different input levels (Smirnakis et al., 1997) but this was not implemented in the present work.

The model provides a phenomenological account of the underlying response reliability. In a given response, the reliability of an input event, i.e., its tendency to fire a spike at each presentation of the same input, is better explained by the difference between the dynamic threshold and the cell's input than by the cell's input alone (Fig. 12). This illustrates that simple integrate-and-fire models provide a convincing phenomenological explanation of spike train statistics, confirming previous work in retinal ganglion cells showing that temporal precision is correlated with the slope of depolarization preceding a spike (Pillow et al., 2005).

Physiological mechanisms of threshold adaptation

In our model, reduced level sensitivity is a consequence of spike threshold adaptation. However, this is only a phenomenological model of the entire early auditory pathway, which was constrained by spikes and not by the membrane potential (which was not recorded). Therefore, we cannot conclude that the measured level sensitivity is due to threshold adaptation. Nevertheless, threshold adaptation is a well-known property of neurons, which has been reported in many areas, both *in vitro* and *in vivo*, in visual cortex (Azouz and Gray, 2000), hippocampus (Henze and Buzsáki, 2001), barrel cortex (Wilent and Contreras, 2005), and in the avian cochlear nucleus (Howard and Rubel, 2010) and inferior colliculus (Peña and Konishi, 2002). This phenomenon was also modeled in several studies (Brandman and Nelson, 2002; Chacron et al., 2003; Kobayashi et al., 2009; Platkiewicz and Brette, 2010, 2011; Brette, 2012).

Spike initiation is modulated by the properties of voltage-gated channels, in particular the inactivation of Nav1.6 channels in the initial segment and the activation of Kv1 channels. These two mechanisms imply that the voltage threshold for spike initiation adapts to the membrane potential (Platkiewicz and Brette, 2010). The resulting dynamic threshold model accounts for threshold properties observed *in vivo* such as the sensitivity to depolarization slope (Platkiewicz and Brette, 2011). In (Brette, 2012), it was shown that under specific conditions, threshold adaptation could produce responses that are insensitive to level. The model we have used in this study, the ATM, is an extension of this level invariant model. The cellular mechanisms responsible for threshold adaptation could be present in the bushy cells of the cochlear nucleus (which are the cells we recorded from), and/or in the AN fibers.

The modeled threshold could also be implemented via a network mechanism. In particular, it was shown in bushy cells of gerbils that the minimum excitatory input required to elicit a spike increases with level (Kuenzel et al., 2011), which is consistent with our model. The authors suggested that this modulation could be due to inhibition tuned at the same preferred frequency, possibly provided by the dorsal cochlear nucleus. This is a possible explanation, but we note that it requires the inhibitory input to be precisely tuned, with the same properties as the bushy cell, not only in frequency tuning (same CF) (Caspary et al., 1994; Gai and Carney, 2008), but with the complete response, including temporal properties. Indeed, the threshold must follow envelope changes occurring at the time scale of the characteristic period of the cell.

Rate of depolarization threshold

An in-vitro study in mice (McGinley and Oertel, 2006) showed that bushy cells have a threshold of rate of depolarization (ROD), i.e., the excitatory input must depolarize the membrane fast enough to trigger a spike (type III excitability). This empirical observation can be reproduced with our adaptive threshold provided that a>1, which is consistent with the values found for the present data (Table 2 and 3). One such realization is shown in Fig. 13. If the ROD of the input is too small (Fig. 13A), the threshold tracks the input and remains above it, so that no spike is triggered. If the ROD is large enough (Fig. 13B), the threshold does not have the time to track the input, which will cross the threshold. Similar to what was measured in vitro (McGinley and Oertel, 2006), there exists a value along the ROD axis - the ROD threshold - above which a cell will always fire (around 7mV/ms for this example). The precise value of the ROD threshold depends on the time constant of the threshold; the faster it is, the larger the ROD threshold.

Limitations of the model and possible extensions

Although the deterministic adaptive model performed very well at predicting spike times at different levels and the stochastic model outperformed the LNP approach, the model could be improved in various ways, at the cost of simplicity. The auditory filtering derived from the reverse correlation, which implements the (linear) filtering of the afferent pathway, was taken here to be constant across levels. For most cells the changes were indeed very small, but a few neural filters showed variations with stimulus level. Time-varying non-linearities controlled by some feedback mechanism could be included, a standard approach used for AN modeling (Tan and Carney, 2003). For instance, the bandwidth of the filters (the inverse of the time constant) (Tan and Carney, 2003) could be a function of input level. The phase of auditory nerve firing for low frequency pure tones is also known to depend on level (Carney and Yin, 1988). More

generally, even at fixed level, the linear front end used in this paper does not capture non-linear stimulus-dependent effects, which is certainly a source of error in the performance of the model.

In a few cells, we observed that spike timing shift was positively correlated with level (increase in absolute spike timing with level) (Fig.4F), which seems paradoxical. In theory, the model cannot reproduce this phenomenon. These paradoxical sensitivities have also been recorded in the AN (Michelet et al., 2012) and at the cochlear level (Recio et al., 1998). Whereas the most plausible mechanisms explaining this counter-intuitive phenomenon are cochlear, cellular mechanisms could also be involved. The subthreshold adaptation of the threshold, which in our model is linear, is in reality non-linear (Platkiewicz and Brette, 2010). This more realistic behavior could be included in the model, and might reproduce the observed paradoxical level sensitivity.

Our initial motivation was to obtain a simple model that can reproduce the spike trains of auditory neurons across a large range of sound levels. In this study, we used broadband stationary noises as acoustical inputs. A logical next step would be to extend the set of sounds to include a variety of ecological sounds. Certainly, predicting the responses of these neurons to a variety of non-stationary sounds will prove challenging.

Acknowledgments

Work supported by a European Community Marie Curie fellowship (PIOF-GA-2011-300753) for B.F., the Fund for Scientific Research - Flanders (G.0714.09 and G.0961.11), and Research Fund K.U.Leuven (OT/05/57 and OT/09/50) for P.X.J. and by the European Research Council (ERC StG 240132) and by Agence Nationale de la Recherche (ANR-11-0001-02 PSL* and ANR-10-LABX-0087) for R.B.

Bibliography

Avissar M, Furman AC, Saunders JC, Parsons TD. Adaptation reduces spike-count reliability, but not spike-timing precision, of auditory nerve responses. *J Neurosci* 27: 6461 –6472, 2007.

Azouz R, Gray CM. Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc Natl Acad Sci U S A* 97: 8110–8115, 2000.

Berry MJ, Warland DK, Meister M. The structure and precision of retinal spike trains. *Proc Natl Aca Sci U.S.A.* 94: 5411–5416, 1997.

de Boer E, de Jongh HR. On cochlear encoding: potentialities and limitations of the reverse-correlation technique. *J Acoust Soc Am* 63: 115–135, 1978.

de Boer E, Nuttall AL. The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions. *J Acoust Soc Am* 101: 3583–3592, 1997.

Brandman R, Nelson ME. A simple model of long-term spike train regularization. *Neural Comput* 14: 1575–1597, 2002.

887 **Brenner N, Bialek W, de Ruyter van Steveninck R**. Adaptive rescaling maximizes information
888 transmission. *Neuron* 26: 695–702, 2000.

889 **Brette R**. Spiking models for level-invariant encoding. *Front Comput Neurosci* 5: 63, 2012.

890 **Carney LH, McDuffy MJ, Shekhter I**. Frequency glides in the impulse responses of auditory-
891 nerve fibers. *J Acoust Soc Am* 105: 2384, 1999.

892 **Carney LH, Yin TC**. Temporal coding of resonances by low-frequency auditory nerve fibers:
893 single-fiber responses and a population model. *J Neurophysiol* 60: 1653–1677, 1988.

894 **Caspary DM, Backoff PM, Finlayson PG, Palombi PS**. Inhibitory inputs modulate discharge
895 rate within frequency receptive fields of anteroventral cochlear nucleus neurons. *J Neurophysiol*
896 72: 2124–2133, 1994.

897 **Chacron MJ, Lindner B, Longtin A**. Threshold fatigue and information transfer. *J Comput*
898 *Neurosci* 23: 301–311, 2007.

899 **Chacron MJ, Pakdaman K, Longtin A**. Interspike interval correlations, memory, adaptation, and
900 refractoriness in a leaky integrate-and-fire model with threshold fatigue. *Neural Comput* 15:
901 253–278, 2003.

902 **Chichilnisky EJ**. A simple white noise analysis of neuronal light responses. *Network* 12: 199–
903 213, 2001.

904 **Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR**. Efficiency and ambiguity in
905 an adaptive neural code. *Nature* 412: 787–792, 2001.

906 **Fischer BJ, Steinberg LJ, Fontaine B, Brette R, Peña JL**. Effect of instantaneous frequency
907 glides on interaural time difference processing by auditory coincidence detectors. *Proc Natl*
908 *Acad Sci U.S.A.* 108: 18138 –18143, 2011.

909 **Fontaine B, Goodman DFM, Benichoux V, Brette R**. Brian hears: online auditory processing
910 using vectorization over channels. *Front Neuroinform* 5: 9, 2011.

911 **Gai Y, Carney LH**. Influence of inhibitory inputs on rate and timing of responses in the
912 anteroventral cochlear nucleus. *J Neurophysiol* 99: 1077–1095, 2008.

913 **Gerstner W, Naud R**. How good are neuron models? *Science* 326: 379–380, 2009.

914 **Gollisch T, Meister M**. Rapid neural coding in the retina with relative spike latencies. *Science*
915 319: 1108–1111, 2008.

916 **Goodman DFM, Brette R**. The Brian simulator. *Front Neurosci* 3, 183-196, 2009.

917 **Hansen N, Ostermeier A**. Completely derandomized self-adaptation in evolution strategies. *Evol*
918 *Comput* 9: 159–195, 2001.

919 **Henze D., Buzsáki G**. Action potential threshold of hippocampal pyramidal cells in vivo is
920 increased by recent spiking activity. *Neuroscience* 105: 121–130, 2001.

921 **Hosoya T, Baccus SA, Meister M**. Dynamic predictive coding by the retina. *Nature* 436: 71–77,
922 2005.

Howard MA, Rubel EW. Dynamic spike thresholds during synaptic integration preserve and enhance temporal response properties in the avian cochlear nucleus. *J Neurosci* 30: 12063 – 12074, 2010.

Irino T, Patterson RD. A compressive gammachirp auditory filter for both physiological and psychophysical data. *J Acoust Soc Am* 109: 2008, 2001.

Jolivet R, Schürmann F, Berger TK, Naud R, Gerstner W, Roth A. The quantitative single-neuron modeling competition. *Biol Cybern* 99: 417–426, 2008.

Jolivet R., Lewis JL, Gerstner W. Generalized integrate-and-fire Models of Neuronal Activity Approximate Spike Trains of a Detailed Model to a High Degree of Accuracy. *J Neurophysiol* 92: 959–976, 2004.

Joris PX, Carney LH, Smith PH, Yin T. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J Neurophysiol* 71: 1022–1036, 1994.

Joris PX, Louage DH, Cardoen L, van der Heijden M. Correlation index: a new metric to quantify temporal coding. *Hear res* 216: 19–30, 2006.

Joris PX, Smith PH. The volley theory and the spherical cell puzzle. *Neuroscience* 154: 65–76, 2008.

Keat J, Reinagel P, Reid RC, Meister M. Predicting every spike: a model for the responses of visual neurons. *Neuron* 30: 803–817, 2001.

Kobayashi R, Tsubo Y, Shinomoto S. Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Front Comput Neurosci* 3: 9, 2009.

Kuenzel T, Borst JGG, van der Heijden M. Factors controlling the input–output relationship of spherical bushy cells in the gerbil cochlear nucleus. *J Neurosci* 31: 4260 –4273, 2011.

Kuhlmann L, Burkitt AN, Paolini A, Clark GM. Summation of spatiotemporal input patterns in leaky integrate-and-fire neurons: application to neurons in the cochlear nucleus receiving converging auditory nerve fiber input. *J Comput Neurosci* 12: 55–73, 2002.

Lichtsteiner P, Posch C, Delbrück T. A 128 X 128 120db 30mw asynchronous vision sensor that responds to relative intensity change. In: *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*., 2006, p. 2060 –2069.

Liu S-C, van Schaik A, Minch BA, Delbrück T. Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms. In: *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*., 2010, p. 2027 –2030.

Louage DHG, van der Heijden M, Joris PX. Enhanced temporal response properties of anteroventral cochlear nucleus neurons to broadband noise. *J Neurosci* 25: 1560 –1570, 2005.

Louage DHG, Joris PX, van der Heijden M. Decorrelation sensitivity of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise. *J Neurosci* 26: 96 – 108, 2006.

Mainen ZF, Sejnowski TJ. Reliability of spike timing in neocortical neurons. *Science* 268: 1503, 1995.

962 **McGinley MJ, Liberman MC, Bal R, Oertel D**. Generating synchrony from the asynchronous:
963 compensation for cochlear traveling wave delays by the dendrites of individual brainstem
964 neurons. *J Neurosci* 32: 9301–9311, 2012.

965 **McGinley MJ, Oertel D**. Rate thresholds determine the precision of temporal integration in
966 principal cells of the ventral cochlear nucleus. *Hear res* 216: 52–63, 2006.

967 **Michelet P, Kovačić D, Joris PX**. Ongoing temporal coding of a stochastic stimulus as a function
968 of intensity: time-intensity trading. *J Neurosci* 32: 9517–9527, 2012.

969 **Nagel KI, Doupe AJ**. Temporal processing and adaptation in the songbird auditory forebrain.
970 *Neuron* 51: 845–859, 2006.

971 **Oertel D**. The role of timing in the brain stem auditory nuclei of vertebrates. *Annu Rev Physiol*
972 61: 497–519, 1999.

973 **Patterson RD**. The sound of a sinusoid: spectral models. *J Acoust Soc Am* 96: 1409–1418,
974 1994.

975 **Peña JL, Konishi M**. From postsynaptic potentials to spikes in the genesis of auditory spatial
976 receptive fields. *J Neurosci* 22: 5652 –5658, 2002.

977 **Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ**. Prediction and decoding of
978 retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci* 25: 11003 –11013,
979 2005.

980 **Platkiewicz J, Brette R**. A Threshold equation for action potential initiation. *PLoS Comput Biol* 6:
981 e1000850, 2010.

982 **Platkiewicz J, Brette R**. Impact of fast sodium channelinactivation on spike threshold dynamics
983 and synaptic integration. *PLoS Comput Biol* 7: e1001129, 2011.

984 **Recio-Spinoso A.** Enhancement and distortion in the temporal representation of sounds in the
985 ventral cochlear nucleus of chinchillas and cats. PLoS One. 7(9):e44286. doi:
986 10.1371/journal.pone.0044286., 2012.

987 **Recio A, Rich NC, Narayan SS, Ruggero MA**. Basilar-membrane responses to clicks at the base
988 of the chinchilla cochlea. *J Acoust Soc Am* 103: 1972–1989, 1998.

989 **Recio-Spinoso A, Narayan SS, Ruggero MA**. Basilar membrane responses to noise at a basal
990 site of the chinchilla cochlea: quasi-linear filtering. *JARO - J Assoc Res Oto* 10: 471–484, 2009.

991 **Rossant C, Fontaine B, Goodman DFM**. Playdoh: A lightweight Python library for distributed
992 computing and optimisation. *J Comput Sci* doi: 10.1016/j.jocs.2011.06.002, in press.

993 **Rossant C, Goodman DFM, Fontaine B, Platkiewicz J, Magnusson AK, Brette R**. Fitting
994 neuron models to spike trains. *Front Neurosci* 5: 9, 2011.

995 **Rossant C, Goodman DFM, Platkiewicz J, Brette R**. Automatic fitting of spiking neuron models
996 to electrophysiological recordings. *Fron Neuroinform* 4: 2, 2010.

997 **Rothman JS, Manis PB**. The roles potassium currents play in regulating the electrical activity of
998 ventral cochlear nucleus neurons. *J Neurophysiol* 89: 3097 –3113, 2003.

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W. Reproducibility and variability in neural spike trains. *Science* 275: 1805–1808, 1997.

Sadeghi SG, Chacron MJ, Taylor MC, Cullen KE. Neural variability, detection thresholds, and information transmission in the vestibular system. *J Neurosci* 27: 771–781, 2007.

Savard M, Krahe R, Chacron MJ. Neural heterogeneities influence envelope and temporal coding at the sensory periphery. *Neuroscience* 172: 270–284, 2011.

Schwartz O, Pillow JW, Rust NC, Simoncelli EP. Spike-triggered neural characterization. *J Vision* 6, 2006.

Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386: 69–73, 1997.

Sumner CJ, Lopez-Poveda EA, O'Mard LP, Meddis R. Adaptation in a revised inner-hair cell model. *J Acoust Soc Am* 113: 893, 2003.

Tan Q, Carney LH. A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. *J Acoust Soc Am* 114: 2007, 2003.

Tollin DJ, Koka K. Postnatal development of sound pressure transformations by the head and pinnae of the cat: monaural characteristics. *J Acoust Soc Am* 125: 980–994, 2009.

Trussell LO. Synaptic mechanisms for coding timing in auditory neurons. *Annu Rev Physiol* 61: 477–496, 1999.

van der Heijden M, Joris PX. Interaural correlation fails to account for detection in a classic binaural task: dynamic ITDs dominate N0Sπ detection. J. Assoc. Res. Otolaryngol. 11:113-131, 2010.

van der Heijden M, Louage DHG, Joris PX. Responses of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise: implications for the sensitivity to interaural delays. J. Assoc. Res. Otolaryngol. 12:485-502, 2011.

Wagner H, Brill S, Kempter R, Carr CE. Auditory responses in the barn owl's nucleus laminaris to clicks: impulse response and signal analysis of neurophonic potential. *J Neurophysiol* 102: 1227–1240, 2009.

Wakeford OS, Robinson DE. Lateralization of tonal stimuli by the cat. *J Acoust Soc Am* 55: 649–652, 1974.

Wilent WB, Contreras D. Stimulus-dependent changes in spike threshold enhance feature selectivity in rat barrel cortex neurons. *J Neurosci* 25: 2983–2991, 2005.

Yin TCT, Chan JCK, Irvine DRF. Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. I. responses to wideband noise. *J Neurophysiol* 55: 280–300, 1986.

Yin TCT, Chan JCK. Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol* 64: 465–488, 1990.

Yin TCT. Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In*: Integrative Functions in the Mammalian Auditory Pathway,* edited by D. Oertel, A. N. Popper, and R. R. Fay. New York: Springer, 2002, p. 99–159.

Young ED, Barta PE. Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J Acoust Soc Am* 79: 426–442, 1986.

Zhang X, Carney LH. Response properties of an integrate-and-fire model that receives subthreshold inputs. *Neural comput* 17: 2571–2601, 2005.

Zhang X, Heinz MG, Bruce IC, Carney LH. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J Acoust Soc Am* 109: 648, 2001.

Zilany MSA, Bruce IC, Nelson PC, Carney LH. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J Acoust Soc Am* 126: 2390, 2009.

Figure captions

Fig.1: Coding of fine-structure in trapezoid body. Dot rasters to 50 presentations of a frozen broadband noise at 6 different stimulus levels (ranging from 40dB SPL up to 90dB SPL). This example is the cell which has been presented the greatest number of input levels in our dataset.

Fig. 2: Schematic illustration of the correlograms calculation. A: Shuffled auto correlogram (SAC) computation. A1: the intervals between a certain spike (surrounded by an ellipse in this example) and every other forward spike in the other trains (but not the spikes from the same train) are computed and tallied in a histogram (Fig. A2). This operation is repeated for every spike of every spike train, the results mirrored (as the auto correlation is symmetric), yielding the SAC (Fig. A3). The half-height width (HHW) is the width of the main lobe where the values are half of the SAC peak. B: Cross stimulus autocorrelogram (XAC) computation. Same as in A but in this case, the intervals are computed between individual spikes from the responses to a repeated stimulus at one stimulus level and all spikes from a repeated stimulus at another stimulus level (including the same trial index). Both forward and backward intervals are considered (Fig. B2). Repeating this operation for every spike yields the XAC (Fig. B3). The lag is defined as the position of the main lobe peak.

Fig.3: High-sync response properties as a function of stimulus level. A-G: The responses of the neurons are characterized using several metrics at different stimulus levels (x-axis). Each point represents a TB fiber response at a given level and data from a single fiber are joined by a solid line. A: Half-height width of the SAC. B: Correlation index. C: mean firing rate. D: intrinsic Γ_{int} with a temporal window of 0.5ms. E: Fano factor. F: lag in µs of the response with respect to a reference (usually at 70dB). G: same as F but with the lag given in cycle (lag in µs*CF). H,I: sensitivity with respect to level of each fiber (defined as the slopes of the linear regressions performed on the curves in (G) and (F)). (H) gives the lag sensitivity in µs/10dB whereas (I) is in cycle/10dB.

Fig.4: Reverse correlation analysis. (A): RevCor examples for a fiber at 6 input levels from 40dB to 90dB (each color represents one level with the darkest color for the lowest level and the brightest color for the highest level) (B): Example of resulting linear gamma chirp fits of the revcors shown in (A). In (C, D, and E) the left column shows resulting fitting parameters as a function of the fiber CF for different input levels. Data points are color-coded, with a dark color for the lower SPLs and a bright color for the higher SPLs. Each line connecting two or more data points represents the data from one fiber (n=24) at multiple SPLs. The right column represents the percentage of change per 10 dB of the corresponding parameter as a function of the fiber CF. This percentage is defined as the level sensitivity (defined as the slope of the linear regression line) divided by the mean value of the parameter. C: width τ of the gamma envelope. D: starting frequency f_0 of the chirp. E: glide slope c in kHz/ms.

Fig. 5: Testing results using a fixed threshold with learning at equal stimulus levels A: The pre-processing performed on the incoming sound consists of the linear filtering implemented as a FIR filter with an impulse response derived from a reverse correlation procedure. The output signal is delayed in order to compensate for synaptic and filtering delays. The delayed signal is then fed to a neuron model. The output spike trains are compared to the target spike trains using a fitness criterion. B: LIF model: after half-wave rectification and compression, the signal is fed to a LIF neuron model which integrates the signal and outputs a spike if it crosses a fixed threshold. The parameters of the LIF model are learned at each stimulus level for each neuron and the testing is performed at the same level, but with an independent input. C: example of spike train responses at two stimulus levels. The rasters in black are the recorded spike trains of the fiber (50 trials of a frozen noise at two levels, 50 dB and 70 dB). The red vertical lines are the output spikes of the LIF neuron model. D: the firing rates of the models (y-axis) against the firing rates of the corresponding fibers at different levels (x-axis). The solid line is the identity diagonal. The data points are color-coded with a dark color for the lower SPLs and a bright color for the higher SPLs. E: Same as in (D) but for the mean gamma factors Γ between the models and the fiber responses (y-axis) against the intrinsic Gamma factors Γ_{int} of the measured responses (x-axis). EV: explained variance, R: coefficient of correlation. F, G: resulting parameters as a function of input level, each line represents a single fiber for the fixed threshold V_T (D), and the compression exponent *c* (E).

Fig.6: Comparison between fixed and adaptive threshold. The spike trains of a fiber recorded at six SPLs were concatenated to train a model but the testing is performed at every level. A: The rasters in black are the measured spikes of a fiber fed with 6 stimulus levels (10ms zoom, 40dB to 90dB in 10dB steps, 50 trials of the same frozen noise at every intensity). The red vertical lines are the resulting spikes of the adaptive threshold model (inset of (E)) and the green ones are those of the fixed threshold model. (A) is a 10-ms chunk of the raster plot. B, C: The corresponding testing results are shown for different stimulus levels as the response lag with respect to a intensity reference (B) and as the mean firing rates (C). The blue curves represent the metrics of the measured data, the red ones are for the adaptive threshold model, and the green ones for the fixed threshold model. The differences in discharge behavior when the input intensity is varied are illustrated in (D) and (E). In (D), it is shown that an increase in intensity

changes the time of firing and the number of spikes generated with a fixed threshold model, whereas a carefully chosen adaptive threshold can yield invariant responses (E). F: The rasters in black are the measured spikes of a fiber fed with 6 stimulus levels (same as in (A) but with a 50ms zoom). The red vertical lines are the resulting spikes of the adaptive threshold model

Fig.7: Effect of previous spikes on firing probability. A: illustration of the procedure used to compute the probabilities $P^i_{0\to1}$ and $P^i_{1\to1}$ (see Material and Methods). For each event, $P^i_{1\to1}$ (resp. $P^i_{0\to1}$) is the probability that a spike is generated at time t (t in event i) given that a spike occurred (resp. did not occur) in a given past temporal window Δ (green box). In order to discard possible effects of the refractory period, for a given event starting at time *t*, spikes occurring in the grey box are discarded. B: Interspike-interval histogram (ISI) for a fiber at one intensity. The refractory period for each fiber is defined as the shortest interval where the ISI histogram exceeds 5% of its maximum (red vertical line). C and D: All of the probabilities paired $(P^i_{0\to1}, P^i_{1\to1})$ are pooled and a 2-dimensional kernel density estimation is performed. Dark blue represent low probabilities and dark red high probabilities. In (C) the time window has a duration extending in the past that equals the fiber characteristic period (CP). In (D) the time window extends from t-2CP to t-CP, where t is the time of the event. In (C) the firing probabilities when a spike occurred in the time window is lower than when no spike occurred, which means that spike history has a suppressive effect on subsequent spikes. This is not the case when the process looks further in the past (D).

Fig.8: Population testing for multi-intensity learning of the deterministic models. For each TB fiber the parameters of the fixed threshold model (row 1), the simple adaptive threshold model (row 2), the level invariant model (row 3), and our adaptive threshold model (row 4) are learned with all stimulus levels available for this fiber at once and testing is performed at every intensity using the corresponding models with independent inputs. The points in the figures, which represent TB fibers at single stimulus levels, are color-coded, with lowest stimulus levels having dark colors and high levels having bright colors. The solid lines are the identity diagonals. A: firing rates of the models (y-axis) against firing rates of the corresponding fibers at different levels (x-axis). B: same as in (A) but for the mean gamma factor between the model and the fiber responses (y-axis) against the intrinsic gamma factor Γ_{int} of the measured responses (x-axis). C: lag of the responses with respect to a reference intensity (usually 70dB) for the measured data (x-axis) and the corresponding models (y-axis). The dashed lines represent the linear regression over the entire dataset. EV: explained variance, R: coefficient of correlation

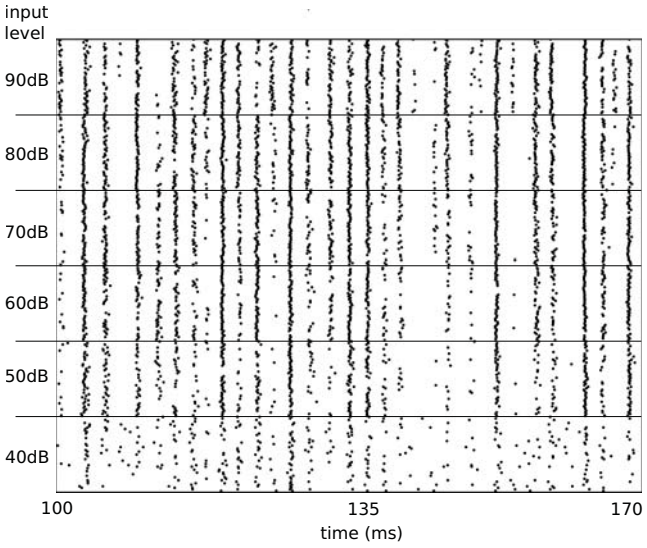
Fig.9: Raster plots of stochastic models learned at equal levels. Responses in the testing set for a recorded TB fiber (A), the corresponding stochastic ATM (B) and LNP model (C), at two different stimulus levels (50dB and 70dB). 50 tokens of the same frozen noise were presented at each intensity. The models were learned at a single intensity and tested at the same intensity with another noise token.

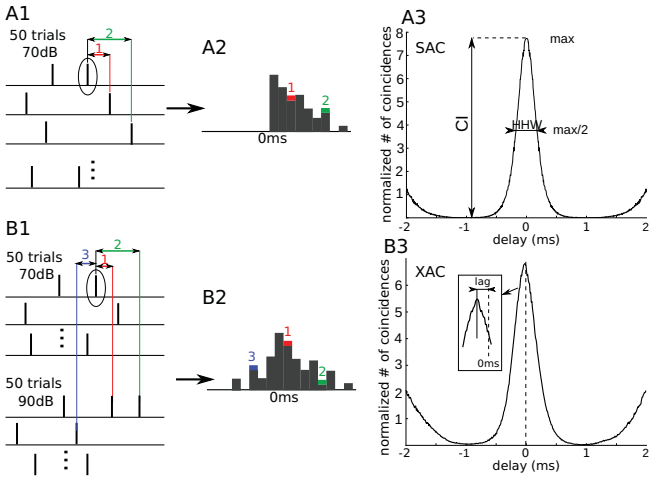
Fig.10: Population results of the stochastic models learned at equal intensity. Row 1 shows the population testing performance of the LNP model, row 2 the population testing performance of the stochastic ATM. In each plot, the model data (y-axis) are plotted against the recorded data (x-axis). Each point, which represents a TB fiber at a given stimulus intensity, is color coded with a dark color for the lower SPLs and a bright color for the higher SPLs. The solid lines are the identity diagonals. (A) compares the firing rates of the models and the data, (B) the HHW, and (C) the correlation index. (D) Correlation coefficients between the two modeled responses (LNP model on x-axis, noisy ATM on y-axis). and the corresponding measured responses. EV: explained variance, R: coefficient of correlation.

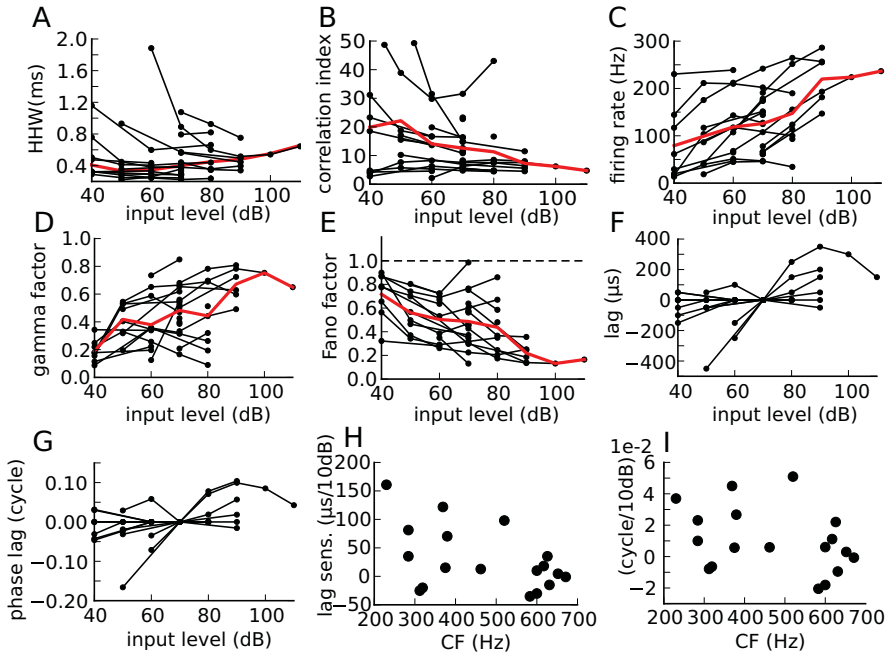
Fig.11: Population tests of the performance of the stochastic ATM for multi-levels learning. For each TB fiber the parameters of the ATM models are learned with all stimulus levels available for this fiber at once and testing is performed at every level using the corresponding models. The points in the figures, which represent TB fibers at single stimulus levels, are color-coded, with lowest stimulus levels having dark colors and high levels having bright colors. The solid lines are the identity diagonals. A: the firing rates of the model responses (y-axis) against the firing rates of the actual responses at different levels (x-axis). B: same as in (A) but for the HHW. C: same as in (A) but for the correlation index. D: lag of the responses with respect to a reference level (usually 70dB) for the measured data (x-axis) and for the corresponding models (y-axis). The dashed line represents the linear regression over the entire dataset. EV: explained variance, R: coefficient of correlation.

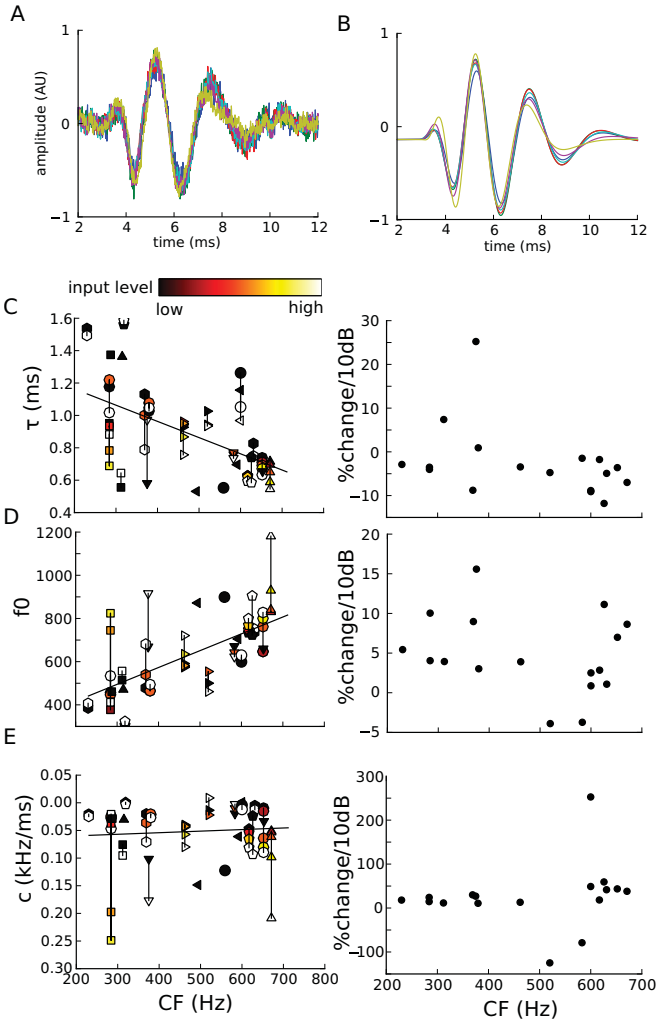
Fig.12: Predicting spike count reliability on the testing set. (A) An event, e.g. E1 or E2, is defined as a time interval where the filtered input stimulus to the modeled neuron is larger than zero. The stochastic thresholds of all the trials (light grey lines) are averaged to yield mean threshold $<V_T(t)>$ (thick black curve). For each event, the distance between stimulus and mean threshold, e.g. d1<0 or d2>0, is defined as the difference between the peak of the stimulus in the event and the value of $<V_T(t)>$ at the beginning of the event. The other panels of (A) show the corresponding raster plot of the 50 presentations of a noise token for the spikes generated by a stochastic ATM (middle row) and for the actual responses (bottom row). In this example the distances d1 and d2 correlate better with the corresponding reliabilities than the peak values of the input in E1 and E2. This procedure is repeated for every event of a TB response. The resulting distance-reliability (B) and peak value-reliability (C) pairs are shown for the same cell. Both distance and peak values are normalized with respect to the mean stimulus level. The thick black line is the fit of an error function to the data points. Each curve yields a coefficient of determination R^2, characterizing the quality of the fit. The R^2 values for the peak values (x-axis) are plotted against the R^2 values for the distances (y-axis) for the responses generated by the stochastic ATM model (D) and the in-vivo recorded responses (E).

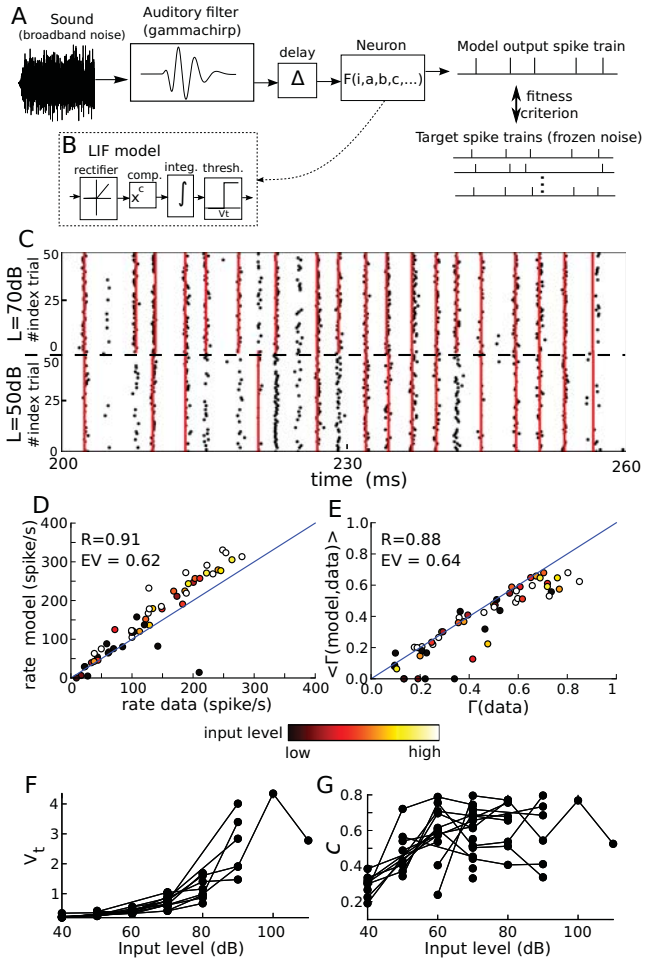
Fig.13: Modeling the rate of depolarization threshold (ROD). If the threshold adapts to the input signal, for slow depolarization (A) the threshold will track the input and no spike will be fired. If depolarization is fast enough (B), the threshold will not have time to track the input and a spike will be fired. . In this example, the step goes from 0.2 mV to 15 mV at different speeds. The relevant model parameters are: a=3, τ=2 ms, α=1, β=1, and r=1 ms.

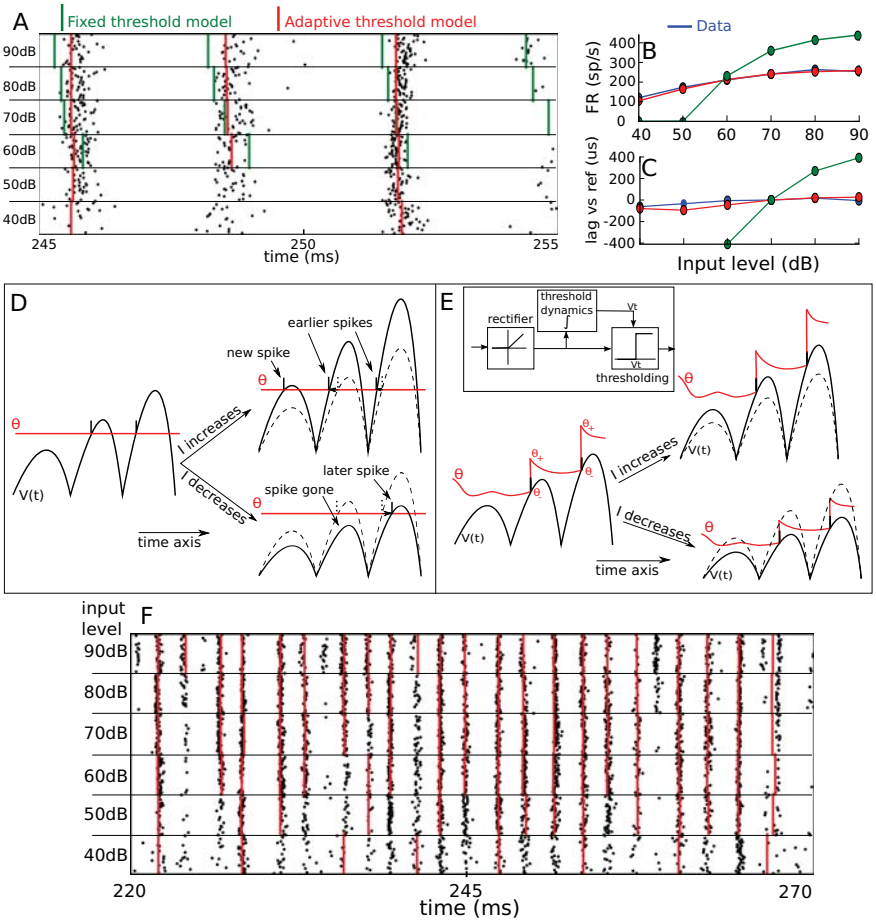

input
level
90dB
80dB
70dB
60dB
50dB
40dB
100
135
170
time (ms)

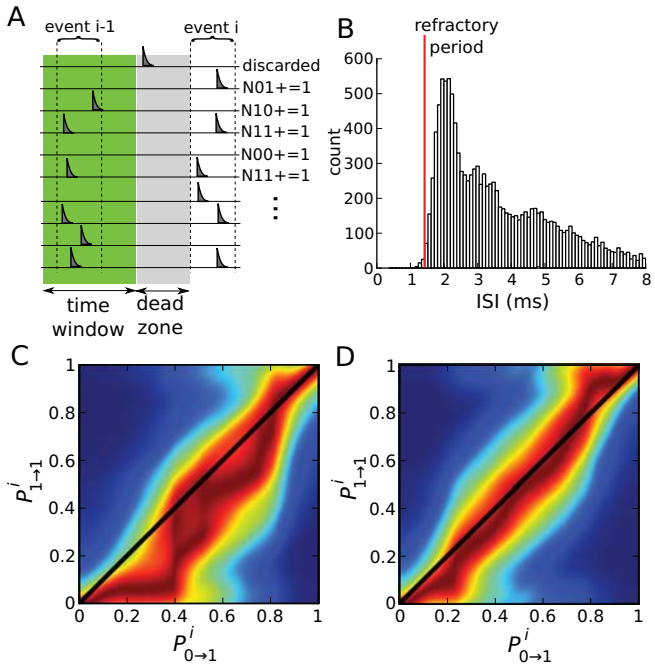

A1
50 trials
70dB
2
1
A2
1
2
0ms
A3
SAC
max
CI
HHW
max/2
normalized # of coincidences
delay (ms)
B1
50 trials
70dB
3
2
1
50 trials
90dB
B2
1
3
2
0ms
B3
XAC
lag
0ms
normalized # of coincidences
delay (ms)

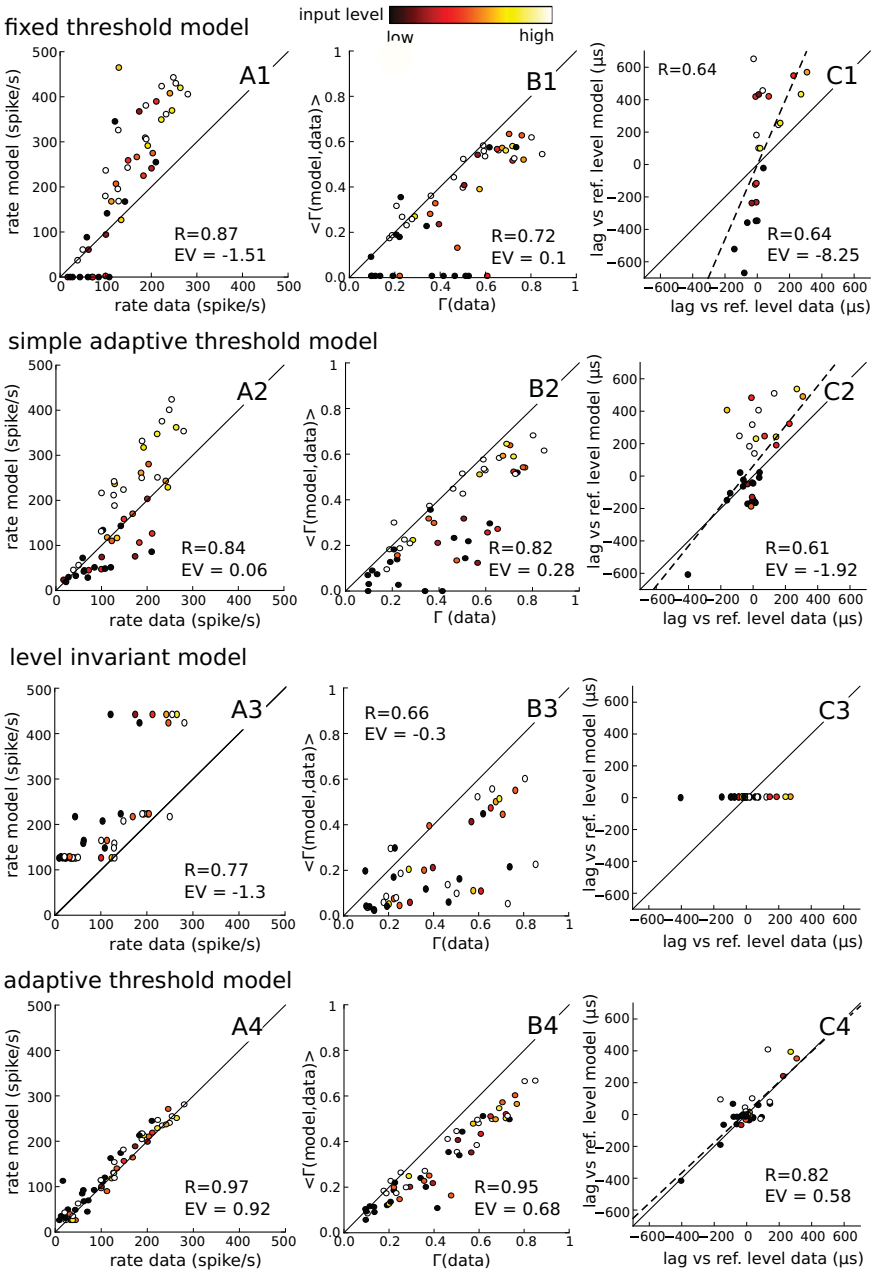

A
HHW(ms)
2.0
1.6
1.2
0.8
0.4
40
60
80
100
input level (dB)
B
correlation index
50
40
30
20
10
0
40
60
80
100
input level (dB)
C
firing rate (Hz)
300
200
100
0
40
60
80
100
input level (dB)
D
gamma factor
1.0
0.8
0.6
0.4
0.2
0.0
40
60
80
100
input level (dB)
E
Fano factor
1.0
0.8
0.6
0.4
0.2
0.0
40
60
80
100
input level (dB)
F
lag (µs)
400
200
0
−200
−400
40
60
80
100
input level (dB)
G
phase lag (cycle)
0.10
0.00
−0.10
−0.20
40
60
80
100
input level (dB)
H
lag sens. (µs/10dB)
200
150
100
50
0
−50
200 300 400 500 600 700
CF (Hz)
I
1e-2
(cycle/10dB)
6
4
2
0
−2
200 300 400 500 600 700
CF (Hz)

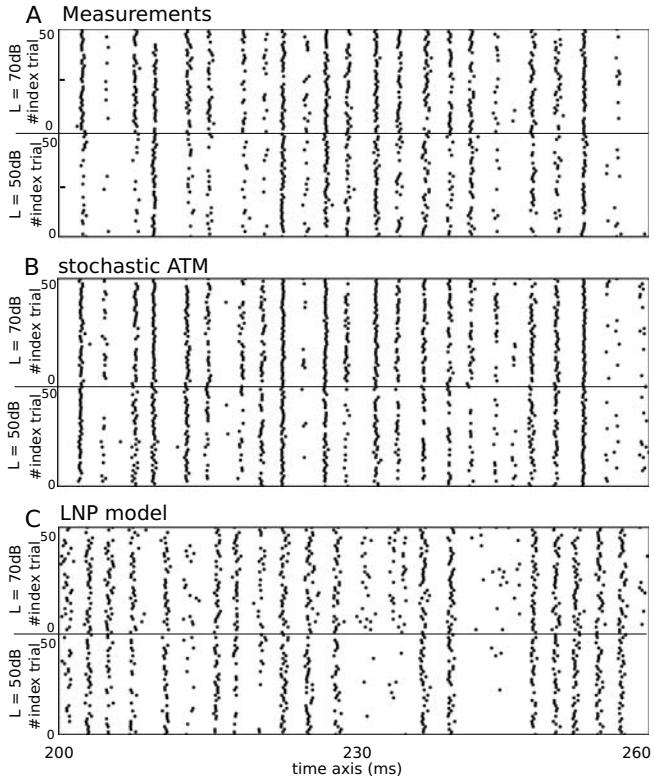

A
B
amplitude (AU)
time (ms)
time (ms)
input level
low
high
C
τ (ms)
%change/10dB
D
f0
%change/10dB
E
c (kHz/ms)
%change/10dB
CF (Hz)
CF (Hz)

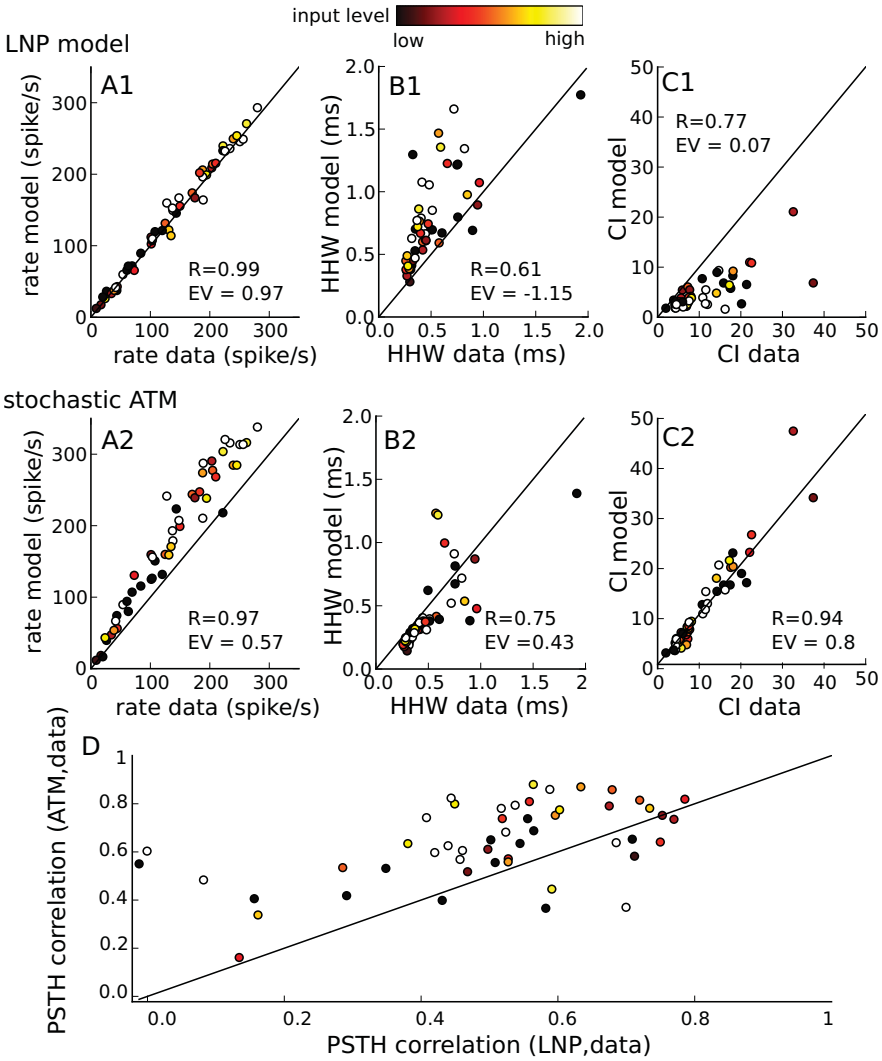

A
Sound
(broadband noise)
Auditory filter
(gammachirp)
delay
Δ
Neuron
F(i,a,b,c,...)
Model output spike train
fitness
criterion
Target spike trains (frozen noise)
B
LIF model
rectifier
comp.
integ.
thresh.
Vt
C
L=70dB
#index trial
L=50dB
#index trial
200
230
260
time (ms)
D
R=0.91
EV = 0.62
rate model (spike/s)
rate data (spike/s)
E
R=0.88
EV = 0.64
<Γ(model,data)>
Γ(data)
input level
low
high
F
Vt
Input level (dB)
G
C
Input level (dB)

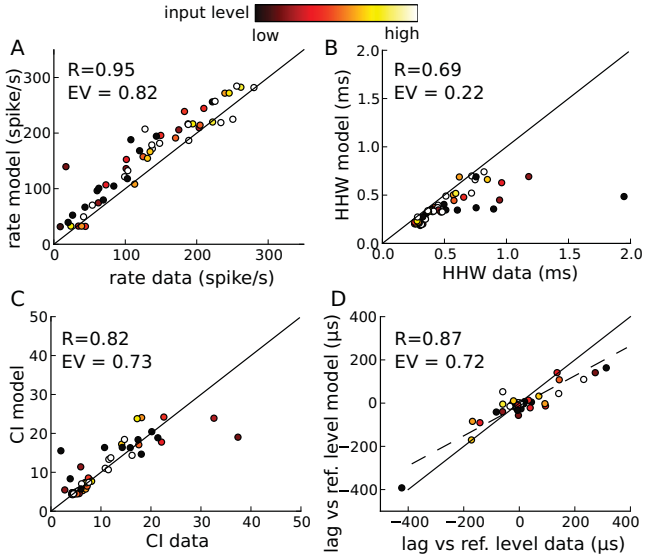

A
Fixed threshold model
Adaptive threshold model
90dB
80dB
70dB
60dB
50dB
40dB
245
250
255
time (ms)
Data
B
FR (sp/s)
400
300
200
100
0
40
50
60
70
80
90
C
lag vs ref (us)
400
200
0
-200
-400
Input level (dB)
D
θ
V(t)
I increases
I decreases
new spike
earlier spikes
spike gone
later spike
time axis
E
rectifier
threshold dynamics
Vt
thresholding
θ+
θ-
I increases
I decreases
V(t)
time axis
input level
F
90dB
80dB
70dB
60dB
50dB
40dB
220
245
270
time (ms)

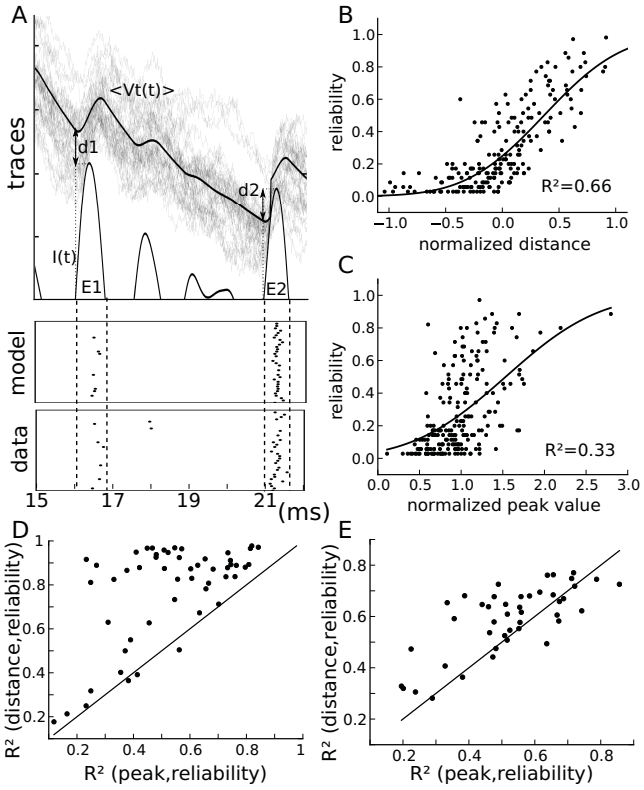

A
event i-1
event i
discarded
N01+=1
N10+=1
N11+=1
N00+=1
N11+=1
time window
dead zone
B
refractory period
count
ISI (ms)
C
D
$P^i_{1\to1}$
$P^i_{0\to1}$

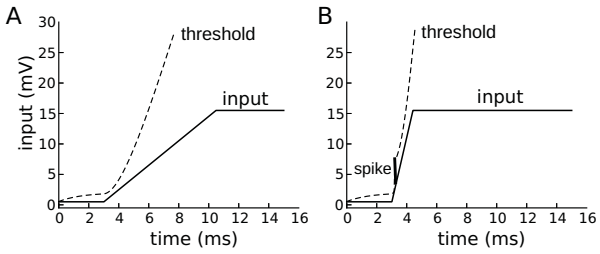

fixed threshold model
input level
low
high
A1
B1
C1
R=0.87
EV = -1.51
R=0.72
EV = 0.1
R=0.64
EV = -8.25
R=0.64
rate model (spike/s)
rate data (spike/s)
<Γ(model,data)>
Γ(data)
lag vs ref. level model (μs)
lag vs ref. level data (μs)
simple adaptive threshold model
A2
B2
C2
R=0.84
EV = 0.06
R=0.82
EV = 0.28
R=0.61
EV = -1.92
Γ (data)
level invariant model
A3
B3
C3
R=0.77
EV = -1.3
R=0.66
EV = -0.3
adaptive threshold model
A4
B4
C4
R=0.97
EV = 0.92
R=0.95
EV = 0.68
R=0.82
EV = 0.58


A
Measurements
L = 70dB
#index trial
50
0
L = 50dB
#index trial
50
0
B
stochastic ATM
L = 70dB
#index trial
50
0
L = 50dB
#index trial
50
0
C
LNP model
L = 70dB
#index trial
50
0
L = 50dB
#index trial
50
0
200
230
260
time axis (ms)


input level
low
high
LNP model
A1
rate model (spike/s)
rate data (spike/s)
R=0.99
EV = 0.97
B1
HHW model (ms)
HHW data (ms)
R=0.61
EV = -1.15
C1
CI model
CI data
R=0.77
EV = 0.07
stochastic ATM
A2
rate model (spike/s)
rate data (spike/s)
R=0.97
EV = 0.57
B2
HHW model (ms)
HHW data (ms)
R=0.75
EV =0.43
C2
CI model
CI data
R=0.94
EV = 0.8
D
PSTH correlation (ATM,data)
PSTH correlation (LNP,data)


input level
low
high
A
R=0.95
EV = 0.82
rate model (spike/s)
rate data (spike/s)
B
R=0.69
EV = 0.22
HHW model (ms)
HHW data (ms)
C
R=0.82
EV = 0.73
CI model
CI data
D
R=0.87
EV = 0.72
lag vs ref. level model (µs)
lag vs ref. level data (µs)


A
traces
<Vt(t)>
d1
d2
I(t)
E1
E2
model
data
15
17
19
21
(ms)
B
reliability
normalized distance
R²=0.66
C
reliability
normalized peak value
R²=0.33
D
R² (distance,reliability)
R² (peak,reliability)
E
R² (distance,reliability)
R² (peak,reliability)


A
B
threshold
input
spike
input (mV)
time (ms)
0
5
10
15
20
25
30
0
2
4
6
8
10
12
14
16