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 and Romain Brette1,2

1Laboratoire Psychologie de la Perception, CNRS, Université Paris Descartes, Paris, France; 2Équipe Audition, Département d’Études Cognitives, Ecole Normale Supérieure, Paris, France; 3Laboratory of Auditory Neurophysiology, KU Leuven, Leuven, Belgium; and 4Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York

Submitted 23 January 2013; accepted in final form 15 July 2013

Fontaine B, Benichoux V, Joris PX, Brette R. Predicting spike timing in highly synchronous auditory neurons at different sound levels. J Neurophysiol 110: 1672–1688, 2013. First published July 17, 2013; doi:10.1152/jn.00051.2013.—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.

bushy cell; level invariance; spike threshold adaptation; temporal coding

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; Joris and Smith 2008; Louage et al. 2005). 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 ITDs over the enormous range of stimulus intensities that the animals experience. Few 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 are 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 sound pressure levels (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 mostly been 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 CN (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. 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 (IF) models (Gerstner and Naud 2009; Jolivet et al. 2008; 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 (Azuaz and Gray 2000; Chacron et al. 2007; Wilent and Contreras 2005). 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 dynamic threshold can successfully reproduce experimental data in visual (Keat et al. 2001), electrosonory (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.

MATERIALS 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 in Louage et al. (2005, 2006) and are only briefly summarized here. Pentobarbital-anaesthetized 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-in. condenser microphone and a probe tube close to the cardrum. 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 SPL in 10-dB steps. Next, a rate-level function was obtained to a broadband Gaussian noise (1 s, repeated every 1.2 s, 5–10 repetitions).

After a fiber’s basic physiological parameters and rate-level functions were collected, a 1-s broadband noise (0.1–30 kHz) with many repetitions was delivered, usually 35–100, to collect at least 3,000 spikes. In some cases, subsequently a second, independent 1-s noise token was similarly delivered. The first input level (overall level 20 μPa) tested was usually 70 dB SPL; the next levels were usually 50, 30, 30, 60, and 10 dB SPL. Because the time we could record from a TB fiber was limited, for certain fibers not all levels are 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 poststimulus time histogram (PSTH) (bin width 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 that 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 RESULTS. It was extensively discussed in previous papers (Joris et al. 1994; Joris and Smith 2008; Louage et al. 2005) that the vast majority, if not all, of these “high-sync” TB fibers are axons of the two varieties of bushy cells: both spherical and globular bushy cells (SBCs and GBCs, respectively). Nevertheless, throughout RESULTS 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. 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) (Joris et al. 2006; Louage et al. 2005). Every possible trial pair is compared (except comparisons of a trial to itself; see 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 (see Fig. 2A2). Since SACs are symmetrical, only forward time intervals are considered. The resulting histogram is then mirrored, yielding the SAC (see Fig. 2A3). The SAC ordinate is normalized by \( n(n−1)Δτ^2D \). This factor eliminates the effect of average rate \( r \), number of presentations \( n \), choice of bin width \( Δτ \), and stimulus duration \( D \). This scaling yields dimensionless bin values. The maximal value of the SAC is referred to as the correlation index (CI; see 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 milliseconds, as half-height width (HHW; see 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 (XACs) 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 (see Fig. 2B1). As XACs are not symmetrical, both forward and backward time intervals are tallied (see Fig. 2, B2 and B3). The XAC \( x \)-axis is normalized by \( n_1n_2Δτr_1r_2D \) where \( n_1 \) and \( n_2 \) are the number of presentations and firing rate of the rth 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 (see Fig. 2B3).

Peripheral model. The model chain, describing the mapping from sounds to spike trains in CN 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 CN and of the neuron itself. It is characterized by an impulse response (see Fig. 5A, auditory filter).

This impulse response is calculated by reverse correlation (Rev-Cor). For a broadband noise stimulus, the RevCor filter \( h(t) \) is the average stimulus that elicits spikes (de Boer and de Jongh 1978; Schwartz et al. 2006), that is,

\[
h(t) = \frac{1}{N} \sum_{n=1}^{N} S(t_n) \]

where \( t_n \) is the time of the nth spike, \( S(t_n) \) is a vector containing the stimulus present in a temporal window preceding that spike, and \( N \) is the total number of spikes in the analysis. On the basis of 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 neuron’s RevCors are first fitted with gamma tone functions (Patterson 1994). The gamma tone is a cosine carrier with a gamma envelope:

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

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

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

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

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

\[
Gc_{log}(t) = A(t - t_0)^3 \exp \left( -\frac{t - t_0}{\tau} \right) \cos\left(2\pi f_0(t - t_0) + 0.5\epsilon(t - t_0)^2 + \theta\right)H(t - t_0)
\]
\[ GC_{\text{lin}}(t) = A(t - t_0)^3 \exp \left( - \frac{t - t_0}{\tau} \right) \cos(2\pi f(t - t_0) + \theta)H(t - t_0) \]

In both \( GC_{\text{lin}} \) and \( GC_{\text{log}} \), the additional parameter \( c \) characterizes the rate of the chirp. As the instantaneous frequency \( f_{\text{inst}} \) is defined as the temporal derivative of the phase, \( f_{\text{inst}} = f_0 + ct \) for \( GC_{\text{lin}} \) and \( f_{\text{inst}} = f_0 + ct \) for \( GC_{\text{log}} \). 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 RevCor in order to minimize the error

\[ x^2 = \frac{1}{N - M} \sum_{i=1}^{N} \left( h(t_i) - g(t_i) \right)^2, \]

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 \( \sigma_i^2 \) 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 decibel, 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 and the fitted parameter value. The slope of this regression is the level sensitivity of the corresponding parameter. This slope is divided by the mean parameter value across levels and multiplied by 100 to yield the percentage of change per decibel.

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

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

\[ \tau_m \frac{dV_m(t)}{dt} = -V_m(t) + I(t) \]

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

\[ \tau_T \frac{dV_T(t)}{dt} = a(I(t) - V_T(t)) \]

where \( V_T \) is the time-varying threshold, \( \tau_T \) is the threshold time constant, and \( a \) quantifies the amount of subthreshold adaptation of \( V_T \). A spike is fired if the input exceeds the threshold: \( I(t) > V_T(t) \). After firing, the threshold is reset: \( V_T \rightarrow \beta V_T + \alpha \), and the neuron cannot fire during a fixed refractory period \( r \). This reset consists of two parts, an additive part \( \alpha \) and a multiplicative part \( \beta \). In Brette (2012), it was proven that a purely multiplicative reset (\( \alpha = 0 \)) yields a level-invariant neuron model, i.e., after a transient time, spike timing and firing rate do not change with stimulus level. To account for some level sensitivity seen in the recorded responses, we include an additive term \( \alpha \). In some cells, adding a second time 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 \( I(t) > V_T(t) + V_T(t) \). 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 \( \alpha = 0, \beta = 0, \) and \( \alpha = 0 \). 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: \( \tau_m, c, V_T, r, \) and \( \Delta \) for the LIF model and \( \Delta, \tau_T, a, \alpha, \beta, \) and \( r \) for the ATM. The model fitting approach employed for the optimization is similar to the one introduced in Rossant et al. (2010, 2011). 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 \( \delta \), the gamma factor \( \Gamma \) (Jolivet et al. 2008):

\[ \Gamma = \left( \frac{2}{1 - 2\delta \exp} \right) \frac{N_{\text{coinc}} - 2N_{\text{exp}} \delta_{\text{exp}}}{N_{\text{exp}} + N_{\text{model}}} \]

where \( \delta_{\text{exp}} \) is the mean firing rate of the experimental response, \( N_{\text{coinc}} \) is the number of coincidences between the model and experimental trains computed within a time window \( \delta \), and \( N_{\text{exp}} \) and \( N_{\text{model}} \) denote the number of spikes in the experiment and model spike train, respectively. \( 2N_{\text{exp}} \delta_{\text{exp}} \) is the expected number of coincidences generated by a Poisson process with rate \( \delta_{\text{exp}} \). The first term in brackets is a normalization factor so that the maximum of \( \Gamma \) 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 \( \delta \). For each cell, there exists a maximum for \( \Gamma \) at a given \( \delta \), \( \delta_{\text{max}} \) (0.5 ± 0.16 ms for the entire population). For the optimization fitness, we set \( \delta = \delta_{\text{max}} \) for each cell. All the optimization results are consistent as long as the chosen \( \delta \) remains in the vicinity of \( \delta_{\text{max}} \), which can be seen as the optimal temporal resolution to compute \( \Gamma \). If they differ too much from \( \delta_{\text{max}} \) (more than a millisecond), the optimizations fail.

The model only outputs one spike train per frozen noise, since it is deterministic, whereas the data contain 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 \( \Gamma \) between the model train and each train of the data, i.e.,

\[ \Gamma_{\text{model, data}}(\Gamma_{\text{model, data}}) = \frac{1}{N} \sum_{n=1}^{N} \Gamma_{\text{i}}, \]

where \( \Gamma_{\text{i}} \) is the gamma factor between the model train and the \( n \)th train out of \( n \) trials. Another useful metric that we will use is the intrinsic gamma factor of a set of repeated trials,

\[ \Gamma_{\text{in}}(\text{data}) = \frac{2}{n(n - 1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \Gamma_{ij} \]

where \( \Gamma_{ij} \) is the gamma factor between trains \( i \) and \( j \) out of \( n \) trials. It quantifies the reproducibility of responses.
We will use a fitness criterion that takes into account the quality both of spike timing prediction and of firing rate prediction:

\[
\text{fitness}(\text{model, data}) = \left[ \frac{\Gamma(\text{model, data}) - \Gamma_{\text{ref}}(\text{data})}{\Gamma_{\text{ref}}(\text{data})} \right] + \lambda \frac{\text{FR}(\text{model}) - \text{FR}(\text{data})}{\text{FR}(\text{data})}
\]

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 \(\lambda = 0.2\) 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 graphics processing units (GPUs) is described in Rossant et al. (2010, 2011). All the neuron simulations and optimizations were performed with 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-s stimulus noise was available, the first 500 ms was used for training and 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-s stimuli were available, the first stimulus was used for learning and the other one for testing. To compute the fitness, the first 50 ms was 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 with 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 with the same learned model. To compute correlations in the recorded data set, spikes from different trials are used, i.e., 50 trials of a 1-s stimulus. Because our models (LIF or ATM) yield identical responses. In particular, differences between firing rates, HHWs, CIs, and the testing is then performed on every single level of the data set, noise is added to the model. This is simply done by adding a white noise term to the threshold equation. The standard deviation of the noise scales linearly with stimulus level, so that the signal-to-noise ratio is constant.

Stochastic adaptive threshold model. To account for stochasticity in the data set, noise is added to the model. This is simply done by adding a white noise term to the threshold equation. The standard deviation of the noise scales linearly with stimulus level, so that the signal-to-noise ratio is constant.

\[
\tau_r \frac{dV_r(t)}{dt} = aI(t) - V_r(t) + \xi(t) \sqrt{2\tau_r \sigma^2(t)}
\]

where \(\xi(t)\) is Gaussian noise, \(\sigma_r\) is the level-independent standard deviation of the noise, and \(I(t)\) is a running average of the input:

\[
\tau_I \frac{dI(t)}{dt} = -I(t) + I(t)\]

We set \(\tau_r\) to 20 ms; \(\sigma_r\) is optimized so that the main lobes of the model response SACs at different levels match the main lobes of the SACs of the corresponding recorded data at the same levels (using a mean square error criterion). When the threshold has two dynamic equations (2 time constants), the same noise is added to both of them.

Linear-nonlinear-Poisson model. We compared our stochastic model with a popular model, the linear-nonlinear-Poisson (LNP) model (Chichilnisky 2001; Pillow et al. 2005). Similar to the two models previously introduced, the input to our LNP model is the filtered sound stimulus. This input is passed through an instantaneous nonlinear function \(f\), which accounts for nonlinearities such as rectification and saturation. The instantaneous spiking probability in response to a stimulus \(s\) is as follows:

\[
P(spike(t)|s) = f(k \ast s(t - \Delta))
\]

where \(s(t)\) is the stimulus, \(\Delta\) is a time delay, \(k\) is the auditory filter impulse response estimated by reverse correlation, \(\ast\) denotes the convolution operator, and \(f\) is a nonlinear function to optimize. Spikes are produced with an inhomogeneous Poisson process. To avoid nonrealistic bursting, a refractory period was set to 1 ms, which is smaller than the shortest characteristic period (1/CF) considered.

Using Bayes’ rule, the nonlinearity function \(f(s) = P(\text{spike}|s)/P(s)\) can be rewritten as \(f(s) = \alpha P(\text{spike}|s)/P(s)\). The prior \(P(s)\) can be estimated with a Gaussian kernel density estimate from the stimulus ensemble. Similarly, \(P(\text{spike})\) was estimated from the spike-triggered stimulus ensemble, i.e., the stimulus values at spike times. \(\alpha\) 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 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. 12A). 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 spiking during this event divided by the total number of trials, \(R_i = p/n\) (Mainen and Sejnowski 1995). An event reliability of 0 means that no spike has been fired, whereas a value of 1 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 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 \((x, y)\) (reliability-distance or reliability-peak magnitude) are fitted, with a least-square method, to the sigmoidal function

\[
f(x) = 0.5 \left(1 + \frac{1}{1 + \exp(-x/x_0)}\right)
\]

where \(w\) is the error function. This is the cumulative distribution function of a normal distribution with mean \(u\) and variance \(\sigma^2\). For each reliability value \(x\), this procedure yields an estimation \(f(x)\) 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 \(R^2 = 1 - \text{SS}_{\text{res}}/\text{SS}_{\text{tot}}\), where the fitted square error \(\text{SS}_{\text{res}} = \sum((y_i - f(x_i))^2)\) and the total empirical variance \(\text{SS}_{\text{tot}} = \sum((y_i - \bar{y})^2)\), with \(\bar{y} = \sum_{i=1}^n y_i\).

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 $t$ (in event $i$) given that a spike occurred in a given past temporal window $\Delta$ (green in Fig. 7A):

$$P_{1 \rightarrow 1} = P(\text{spike at } t \in \text{event } i \mid \text{spike in } \Delta) = \frac{N_{11}}{N_{10} + N_{11}}$$

where $N_{11}$ is the number of trials in which there was a spike both in the preceding window $\Delta = [t-r, t-t_0]$ and in event $i$ ($t$ represents the beginning of the event; green box in Fig. 7A). $N_{10}$ is the number of trials in which there was a spike in the preceding window $\Delta$ 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 $\Delta$:

$$P_{0 \rightarrow 0} = P(\text{spike at } t \in \text{event } i \mid \text{no spike in } \Delta) = \frac{N_{01}}{N_{00} + N_{01}}$$

$N_{01}$ is the number of trials in which no spike occurred in $\Delta$ but a spike occurred in event $i$, and $N_{00}$ is the number of trials in which there was a spike in neither $\Delta$ nor event $i$. For a given event, if spikes that occurred in the past in $\Delta$ have a suppressive effect on subsequent spikes, then $P_{0 \rightarrow 0} > P_{1 \rightarrow 1}$. 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 interspike interval histogram exceeds 5% of its maximum (see Fig. 7B). This procedure is repeated for every event of every response. For visualization (see Fig. 7, C and D), all points ($P_{0 \rightarrow 1}, P_{1 \rightarrow 1}$) are used to estimate the joint probability density via a twodimensional 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 CN 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 $< 1,000$ Hz) with at least one level for which the responses are reliable enough (CI $> 5$). All these cells were classified as PHL. Given the PSTD 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 at 2 SPLs, and 10 at $> 3$ SPLs. Stimulus levels were separated by at least 20 dB. For all cells at least one 1-s frozen noise was presented between 30 and 50 times. When time allowed ($n = 10$), another 1-s 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 MATERIALS AND METHODS), we estimate the temporal precision (Fig. 3A) and the 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: 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 CIs tend to decrease with increasing level. Figure 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 $~3,000$. Thus the firing rate responses exhibit strong compression.

Another standard measure of the temporal precision of the responses is the intrinsic gamma factor $\Gamma_{int}$ (Jolivet et al. 2008). From visual inspection of the raster plots (Jolivet et al. 2008), we define the following selection criteria: we select low-frequency cells (CF $< 1,000$ Hz) with at least one level for which the responses are reliable enough (CI $> 5$). All these cells were classified as PHL. Given the PSTD 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 at 2 SPLs, and 10 at $> 3$ SPLs. Stimulus levels were separated by at least 20 dB. For all cells at least one 1-s frozen noise was presented between 30 and 50 times. When time allowed ($n = 10$), another 1-s 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 MATERIALS AND METHODS), we estimate the temporal precision (Fig. 3A) and the 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: 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 CIs tend to decrease with increasing level. Figure 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 $~3,000$. Thus the firing rate responses exhibit strong compression.
2008), which quantifies the coincidences between responses across trials, at a given temporal resolution \( \delta \). \( \Gamma_{\text{int}} = 0 \) means that there are no more coincidences between trials than expected by chance for a Poisson process, and \( \Gamma_{\text{int}} = 1 \) means that all the trials are identical at the given time resolution. Figure 3D shows \( \Gamma_{\text{int}} \) for all cells and levels, computed at resolution \( \delta = 0.5 \text{ ms} \) 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 data set (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: Avissar et al. 2007; Young and Barta 1986; 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 (ILDs) (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 MATERIALS 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 \( \mu \text{s} \), while stimulus level varies by \( >40 \text{ dB} \). 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 microseconds per 10 dB and in Fig. 3I in cycles per 10 dB. It appears that spike timing in these cells is not very sensitive to input stimulus level (median \( = 16 \mu \text{s}/10 \text{ dB} \) in Fig. 3H, 0.9 \( \times 10^{-2} \) cycles/10 dB in Fig. 3I) when considering the natural range of ILDs. Indeed, at these frequencies, the maximum ILD is \( \sim 5 \text{ dB} \) (Tollin and Koka 2009), yielding a median lag of 8 \( \mu \text{s} \) (see DISCUSSION).

Reverse correlation analysis. We estimated the reverse correlation filters of the neurons from responses to broadband noises, using a 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 \( \pm 0.02 \)), 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 nonchirping (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 CN of barn owls (Fischer et al. 2011; Wagner et al. 2009) or in the AN of cats (Carney et al. 1999). We found no significant difference in estimation error between the logarithmic and the linear chirp functions (median linear chirp \( \chi^2 = 0.043 \), median logarithmic chirp \( \chi^2 = 0.044 \), \( P = 0.42 \), Mann-Whitney U-test). Figure 4B

\[ \text{Fig. 3. High-sync response properties as a function of stimulus level. A–G: responses of the neurons are characterized with several metrics at different stimulus levels (x-axis). Each point represents a TB fiber response at a given level, and data from a single fiber are joined by a solid line. A: HHW of the SAC. B: correlation index. C: mean firing rate. D: intrinsic } \Gamma_{\text{int}} \text{ with a temporal window of 0.5 ms. E: Fano factor. F: lag in } \mu \text{s} \text{ of the response with respect to a reference (usually at 70 dB). G: same as F but with the lag given in cycles [lag in } \mu \text{s} \times \text{characteristic frequency (CF)]}. H \text{ and I: sensitivity with respect to level of each fiber (defined as the slopes of the linear regressions performed on the curves in G and F). H gives the lag sensitivity in } \mu \text{s}/10 \text{ dB, whereas I is in cycles/10 dB.} \]
shows the linear gammachirp 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–E). Figure 4, C–E, left, show the envelope widths \( \tau \) (Fig. 4C), starting frequencies of the chirps \( f_0 \) (Fig. 4D), and gliding slopes \( c \) (Fig. 4E), for every cell and every level as a function of their CF. Each cell is represented by a shape, 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 \( \tau \) is inversely correlated with CF (mean: 0.88 ms, regression \( -10^{-4} \) ms/Hz x +1.36 ms, \( r = -0.55, P < 10^{-5} \)). The starting frequency \( f_0 \) is small and around zero (mean % of change/10 dB = 3 ± 19%).

The sensitivity of the gliding slope is higher because of one outlier (mean % of change/10 dB = 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 \( \theta \) is small and around zero (mean % of change/10 dB = 3 ± 19%).

From these observations, we can conclude that the stimulus level has little effect on the shape of the RevCor. 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 \( n = 16 \) and 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. 5A). Figure 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 = 462 Hz, firing rate = 180 Hz for 50 dB and 240 Hz for 70 dB). 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. 5, D and E). Each situation corresponds to a specific set of parameter values. Figure 5D shows a very good agreement between the model firing rate and the firing rate of the data (correlation coefficient \( R = 0.91, EV = 0.62 \)). Figure 5E shows that spike timing is also well predicted: the gamma factor between model spike trains and data spike trains \( \Gamma \) (model, data) is close to the intrinsic gamma factor of the data \( \Gamma_{\text{in}} \) (data) \( (R = 0.88, EV = 0.64) \). The statistics of the resulting parameters are shown in Table 1 for all cells at all levels.
As we previously noted, there is one set of parameter values for each level. In particular, the fitted values depend on level. Figure 5, F 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 six 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. 5: to correctly predict responses across levels, the spike threshold must increase with level. Figure 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 subthreshold events may become suprathreshold 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 ATM (Fig. 6E, inset) recently proposed in Brette (2012) that has level-invariant responses, in terms of both 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. To take into account deviation from complete level invariance, an additive term is added to the reset (see MATERIALS AND METHODS).

When the same single set parameter optimization procedure is applied to this adaptive model, the prediction performance across levels drastically improves (Fig. 6, A and F, red line). 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 \( P_{1\rightarrow t} \) that a spike is generated at time \( t \) (in event \( i \)) given that a spike occurred in a given past temporal window \( \Delta \) (green box in Fig. 7A). Second, we compute the probability \( P_{0\rightarrow t} \) for the \( i \)th event that a spike is
generated given that no spike previously occurred in $\Delta$. For a given event $i$, if spikes that occurred in the past in $\Delta$ have a suppressive effect on subsequent spikes, then $P'_{0 \rightarrow 1} > P'_{1 \rightarrow 1}$. To discard possible effects of the absolute refractory period, for a given event starting at time $t$, spikes that 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 interspike interval histogram exceeds 5% of its maximum (Fig. 7B), yielding $1.4 \pm 0.4$ ms for the population. This procedure is repeated for every event of every response, and we plot the two-dimensional density $P'_{1 \rightarrow 1}, P'_{0 \rightarrow 1}$ for two different time windows $\Delta$. The first window is $\Delta = [t, t - CP]$, where CP is the characteristic period of the neuron and $t$ is the starting point of the event (Fig. 7C). The second window is one period earlier: $\Delta = [t - CP, t - 2CP]$ (Fig. 7D). We can see in Fig. 7C that $P'_{1 \rightarrow 1}$ is significantly lower than $P'_{0 \rightarrow 1}$ (most of $P'_{1 \rightarrow 1}$ 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 ~1 CP, and that this effect is not due to the refractory period.

**Population analysis of multiple-levels models.** Figure 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. Figure 8, A1–C1, show 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 a lower firing rate at low levels (Fig. 8A1, 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 70 dB) 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 ($R = 0.64$, $EV = -8.25$).
results (Fig. 8, A4–C4) 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; \( R = 0.95 \), \( EV = 0.68 \)). Figure 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.

Stochastic models. So far, we have 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 \( \sigma^2 \) to the spike threshold (see MATERIALS AND METHODS). To maintain a constant signal-to-noise ratio, the standard deviation \( \sigma = \sigma I \) where \( I \) is a low-pass-filtered version of the input \( I \). The invariant part \( \sigma \) of the noise variance is optimized so that the main lobes of the SACs of the model responses at different levels match the main lobes 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 LNP model (Chichilnisky 2001; Pillow et al. 2005). The LNP model consists of a cascade of a linear and a nonlinear stage, followed by Poisson spike generation (see MATERIALS AND METHODS). The linear part is the same auditory filter as previously used, whereas the static nonlinearity is optimized on the learning data set. 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 (Fig. 9A), the stochastic ATM (Fig. 9B), and the optimized LNP model (Fig. 9C). 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 is 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. 10, 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 low (Fig. 10B1; \( R = 0.61 \) and \( EV = -1.15 \)), and the CI is too low (Fig. 10C1; \( R = 0.77 \) 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. Figure 10D shows the correlation for the ATM against the correlation for the LNP model for all cells and levels. The PSTHs are clearly better predicted by our adaptive model than by the LNP model (2-sided t-test: \( P = 6 \times 10^{-8} \), 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 at all, and precision decreases (HHW increases) with increasing level. For the stochastic ATM, the prediction performance for the firing rate is shown in Fig. 11A. Although the model slightly overestimates the firing rate, its predictions are good across the entire level range (\( R = 0.95, EV = 0.82 \)). The predicted temporal precision is also slightly higher than the precision of the cells (HHWs are lower for the models than for the recorded data), but they are good on average (Fig. 11B; \( R = 0.69, EV = 0.22 \)). The CI is also well predicted (Fig. 11C; \( R = 0.82, EV = 0.73 \)). Finally, the prediction performance on the response lags is also very high (Fig. 11D; \( R = 0.87, EV = 0.72 \)). The linear regression (Fig. 11D, dashed line) suggests that the model is in general not sensitive enough to stimulus level.
Predicting spike reliability. It can be seen in Fig. 9A that for a given stimulus some spiking events are more reliable than others. By “reliable event,” we mean that spikes are observed in most trials in the corresponding event, which is defined as an interval where the filtered stimulus \( \tilde{l}(t) \) is positive (e.g., events \( E1 \) and \( E2 \) in Fig. 12A). For each event, we define reliability \( R \) in a similar way as Mainen and Sejnowski (1995), as the proportion of trials in which the neuron spikes in the event. The total number of trials ranges between 35 and 100 in our data set. \( R = 0 \) means that no spike was observed in response to the stimulus event (unreliable event), whereas \( R = 1 \) means that a spike was observed in every trial, i.e., the response is perfectly reliable.

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

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

We then compare the prediction performance of the two quantities, distance and peak, on the entire population (all cells, all levels). For every response and every stimulus level, we calculate the reliability, distance, and peak value for all events, and we fit sigmoid functions to the resulting sets of points (1 set for reliability vs. distance, 1 set for reliability vs. peak), as in Fig. 12, B and 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 data set, 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 modeled 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 have presented a phenomenological model of the responses of CN neurons to broadband sounds. It consists of a linear filter followed by an IF 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 \( \sim 5 \) dB (Tollin and Koka 2009). Given that the population median sensitivity to input level is \( 16 \mu s/10 \) dB (Fig. 3H), the maximal change in timing across the two ears, expected from ILDs, is \( \sim 8 \mu s \) (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, because of the paucity of MOSO 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 MOSO 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 detail. 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 AN response (Sumner et al. 2003; Zhang et al. 2001; Zilany et al. 2009) and of cellular models of neurons of the CN (Kuhlmann et al. 2002; McGinley et al. 2012; Rothman and Manis 2003; Zhang and Carney 2003; Zhang et al. 2001; Zilany et al. 2009) and of cellular sources of noise. Similar to previous results (Pillow et al. 2005), we found that a noisy IF 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 IF 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 the avian CN (Howard and Rubel 2010) and inferior colliculus (Peña and Konishi 2002). This phenomenon was also modeled in several studies (Brandman and Nelson 2002; Brette et al. 2012; 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 Fig. 8). The motivation for the inclusion of a compression on the input in the LIF was twofold: 1) to give a fair chance to the IF model, which would be immediately discarded in the absence of any compression, and 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 for both the firing rate and the precise timing of spikes (Fig. 6 and Fig. 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 neuroengineering 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 IF 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.
nisms responsible for threshold adaptation could be present in the bushy cells of the CN (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 CN. 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 CP 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 (Tables 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 (~7 mV/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 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 nonlinearities 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)
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 nonstationary sounds. Certainly, predicting the responses of these neurons to a variety of nonstationary sounds will prove challenging.

GRANTS

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

Author contributions: B.F. and R.B. conception and design of research; B.F. and P.X.J. performed experiments; B.F., analyzed data; B.F., V.B., and R.B. interpreted results of experiments; B.F. prepared figures; B.F. and R.B. drafted manuscript; B.F., P.X.J., and R.B. edited and revised manuscript; B.F., V.B., P.X.J., and R.B. approved final version of manuscript.

REFERENCES


*J Neurophysiol* • doi:10.1152/jn.00051.2013 • www.jn.org