Emergence of band-pass filtering through adaptive spiking in the owl’s cochlear nucleus

Bertrand Fontaine,1 Katrina M. MacLeod,2 Susan T. Lubejko,2 Louisa J. Steinberg,1 Christine Köppl,3 and Jose L. Peña1

1Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York; 2Department of Biology, Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland; and 3Cluster of Excellence “Hearing4All” and Research Center Neurosensoric Science and Department of Neuroscience School of Medicine and Health Science, Carl von Ossietzky University, Oldenburg, Germany

Submitted 14 February 2014; accepted in final form 27 April 2014

Fontaine B, MacLeod KM, Lubejko ST, Steinberg LJ, Köppl C, Peña JL. Emergence of band-pass filtering through adaptive spiking in the owl’s cochlear nucleus. J Neurophysiol 112: 430–445, 2014. First published April 30, 2014; doi:10.1152/jn.00132.2014.—In the visual, auditory, and electrosensory modalities, stimuli are defined by first- and second-order attributes. The fast time-pressure signal of a sound, a first-order attribute, is important, for instance, in sound localization and pitch perception, while its slow amplitude-modulated envelope, a second-order attribute, can be used for sound recognition. Ascending the auditory pathway from ear to midbrain, neurons increasingly show a preference for the envelope and are most sensitive to particular envelope modulation frequencies, a tuning considered important for encoding sound identity. The level at which this tuning property emerges along the pathway varies across species, and the mechanism of how this occurs is a matter of debate. In this paper, we target the transition between auditory nerve fibers and the cochlear nucleus magnocellularis (NA). While the owl’s auditory nerve fibers simultaneously encode the fast and slow attributes of a sound, one synapse further, NA neurons encode the envelope more efficiently than the auditory nerve. Using in vivo and in vitro electrophysiology and computational analysis, we show that a single-cell mechanism inducing spike threshold adaptation can explain the difference in neural filtering between the two areas. We show that spike threshold adaptation can explain the increased selectivity to modulation frequency, as input level increases in NA. These results demonstrate that a spike generation nonlinearity can modulate the tuning to second-order stimulus features, without invoking network or synaptic mechanisms.

band-pass filtering; envelope encoding; cochlear nucleus; threshold adaptation

Sensory systems have evolved to efficiently represent the statistics of natural stimuli (Barlow 1961). In the auditory system, the cochlea decomposes sound into narrow frequency channels. The output of each cochlear channel can be characterized by its first-order attribute, the time-dependent pressure wave or fine-structure, and second-order attribute, the instantaneous amplitude or envelope (Attias and Schreiner 1997; Lewicki 2002). The fine structure is used, for instance, in sound localization (Yin 2002) and pitch perception (Winter 2005), whereas the envelope is important for sound identification (Nelson and Takahashi 2010; Schneider and Woolley 2010; Singh and Theunissen 2003), including speech perception (Shannon et al. 1998). Downstream the central nervous system, neural pathways become specialized for encoding specific stimulus features. In the barn owl, the auditory stream is segregated one synapse after the cochlea into pathways that encode level and timing via nucleus angularis (NA) and the nucleus magnocellularis (NM), respectively (Köppl 1997; Sullivan and Konishi 1984). Recently, it has been shown that NA neurons encode the sound envelope more reliably than NM neurons (Steinberg and Peña 2011), suggesting that the timing and level pathways may be specialized in encoding first- and second-order features of sound, respectively.

Along the ascending auditory pathway, neurons become increasingly selective for the frequency of the envelope modulation (Frisina 2001; Joris 2004; Krishna and Semple 2000; Rodríguez et al. 2010b; Woolley and Casseday 2005). This selectivity is the result of band-pass filtering on the sound envelope, i.e., neurons respond preferably to certain modulation frequencies. Having neurons heterogeneously tuned to modulation frequency may be important for encoding complex natural stimuli (Woolley et al. 2005). For instance, the higher auditory system of songbird contains neurons that respond selectively to learned songs, suggesting a feature detecting strategy (Gentner and Margoliash 2003). While it is clear that the auditory system must efficiently filter the stimulus envelope, the underlying neural mechanism in the early auditory system, either at the cellular or circuit levels, is still not fully understood.

Neural filtering properties can emerge by different mechanisms: network architecture, e.g., interplay of excitation and inhibition (Nelson and Carney 2004); synaptic plasticity, e.g., short-term synaptic plasticity (Fortune and Rose 2001); resonant membrane conductances (Hutcheon and Yarom 2000); and adaptation (Benda et al. 2005, 2010; Hewitt and Meddis 1993; Higgs and Spain 2011; Platkiewicz and Brette 2011). Filtering properties are dynamic in that they adapt to stimulus statistics such as mean or variance in neurons of the visual (Fairhall et al. 2001) and auditory (Dean et al. 2008; Lesica and Grothe 2008a; Nagel and Doupe 2006; Wen et al. 2009) systems. In the auditory system, changes in stimulus intensity can evoke changes in both spectral (preferred acoustic frequency) and temporal (preferred modulation frequency) processing (Frisina 2001; Krishna and Semple 2000; Lesica and Grothe 2008a; Nagel and Doupe 2006). Neural adaptation to input statistics has important consequences regarding coding efficiency (Dean et al. 2008; Lesica and Grothe 2008b; Rodrí-
from 0.5 to 12 kHz.
imperfections in the amplitude and phase response of each earphone
ition software then used these calibration data to automatically correct
earphones at the beginning of each experiment while inside the owl's
Earphones were constructed from a small speaker (Knowles 1914, Custom software was used to generate stimuli and collect data. double-walled sound-attenuating chamber (Industrial Acoustics). repeated every 7–10 days for a period of several weeks.
also received an intramuscular injection of prophylactic antibiotics (oxytetracycline, 20 mg/kg; Phoenix Pharmaceuticals) and a subcutaneous injection of lactated Ringer solution (10 ml) at the beginning for each experiment. Body temperature was maintained throughout the experiment with a heating pad (American Medical Systems) or held at 39°C by a feedback-controlled blanket system (Harvard Instruments). A metal piece was fixed to the skull with dental cement, to hold the head firmly. A small steel post was also implanted to demarcate a reference point for stereotaxic coordinates. Subsequently, all recording sessions were performed while the head was held in place by the head plate.
For ANF, the bone and meninges overlying the right cerebellum were removed, and the posterior part of the right cerebellum was aspirated to expose the surface of the auditory brain stem on that side. These experiments were terminal. For NA, a craniotomy was performed at the coordinates for the recording site, and a small incision was made in the dura mater for electrode insertion. At the end of a recording session, the craniotomy was sealed with Rolyan silicone elastomer (Sammons Preston). After the experiment, analgesics (ketoprofen, 10 mg/kg. Ketofen; Merial) were administered. Depending on the owl’s weight and recovery conditions, experiments were repeated every 7–10 days for a period of several weeks.

Acoustic stimulation. All in vivo recordings were performed in a double-walled sound-attenuating chamber (Industrial Acoustics). Custom software was used to generate stimuli and collect data. Earphones were constructed from a small speaker (Knowles 1914, Aiwa HP-V14 or Etymotic ER-2) and a microphone (Knowles 1319 or Knowles FG 23329) in a custom-made case that fits the owl’s ear canal. The microphones were calibrated using a Bruel and Kjaer microphone. The calibrated microphones were used to calibrate the earphones at the beginning of each experiment while inside the owl’s ear canal. The calibration data contained the amplitudes and phase angles measured in frequency steps of 100 Hz. The stimulus generation software then used these calibration data to automatically correct irregularities in the amplitude and phase response of each earphone from 0.5 to 12 kHz.

In vivo electrophysiology. For ANF, glass microelectrodes, filled with either 2 M Na-acetate or 3 M KCl and with impedances mostly between 30 and 60 MΩ, were positioned under visual control above the surface of the brain stem and then advanced under remote control. Neural signals were serially amplified by a WPI 767 (World Precision Instruments) and an AC amplifier (PC1, Tucker-Davis Technologies). NA units were recorded using 5 MΩ tungsten electrodes (A-M Systems) and amplified by a DP-301 differential amplifier (Warner Instruments). For NA, brain stem structures were targeted by known stereotaxic coordinates. NA was approached at a 5–10° angle in the coronal plane (Steinberg and Peña 2011). The recording site was determined by first locating the nucleus laminaris and tilting the electrode laterally, until neural responses to only the ipsilateral side were found. NA units can be distinguished from other short-latency monaural-responding structures, namely NM and the auditory nerve, by their poor phase locking, which was measured online and confirmed post. For both ANF and NA recordings, a spike discriminator (SD1, Tucker-Davis Technologies) converted neural impulses into transistor-transistor logic pulses for an event timer (ET1, Tucker-Davis Technologies), which recorded the timing of the pulses. In parallel, the analog waveforms were stored in a personal computer via an analog-to-digital converter (DD1, Tucker Davis Technologies) with a sampling rate of 48 kHz and 16-bit resolution.

In vivo data collection. All stimulation protocols were exclusively ipsilateral. Data used to compute coherence, to analyze the reliability of neuronal responses, and to fit the models were obtained using a frozen noise protocol, in which a single 200-ms (500-ms) white noise stimulus (with power between 0.5 and 12 kHz) was repeatedly presented to ANF (or NA). Data for reverse correlation were obtained by presenting 200-ms (500-ms) de novo-synthesized broadband signals for ANF (or NA). Stimuli were presented with an interstimulus interval of 500 ms, and we attempted to collect over 4,000 spikes for each stimulus condition. For ANF, three sound levels were randomly presented within a range from near threshold to saturation levels. For NA, generally more levels were used (between 2 and 7), ranging from 30 dB up to 90 dB. For some NA cells, a ramped (150 ms up and 450 ms down) 90-dB noise was presented.

In vitro electrophysiology in NA. In vitro physiological experiments were performed as reported previously (Kreeger et al. 2012) and are, therefore, briefly described below. All animal procedures were performed with the approval of the University of Maryland Institutional Animal Care and Use Committee and followed NIH guidelines on animal welfare. Chicken (Gallus gallus domesticus, Charles River SPAFAS, Lebanon, CT) embryos aged 17–18 days incubation (E17–18) were cooled and rapidly decapitated, and the head section containing the brain stem was submerged in chilled and oxygenated low sodium artificial cerebral spinal fluid (ACSF) (in mM: 97.5 NaCl, 3 KCl, 2.5 MgCl2, 26 NaHCO3, 2 CaCl2, 1.25 NaH2PO4, 10 dextrose, 3 HEPES, 230 sucrose). Transverse slices (250 µm thick) through the brain stem containing NA were cut on a vibrating tissue slicer (Leica Microsystems, Wetzlar, Germany) and incubated in oxygenated normal ACSF (in mM: 130 NaCl, 3 KCl, 2 MgCl2, 26 NaHCO3, 2 CaCl2, 1.25 NaH2PO4, 10 dextrose, 3 HEPES) for 30 min at 34°C. Slices were held in ACSF at room temperature until recording, during which they were submerged and continuously perfused with warmed (29.6 ± 1.4°C), oxygenated normal ACSF (1–2 ml/min) in the recording chamber (Series 20 Warner Instruments). No synaptic channel blockers were used. It has been demonstrated (Fukui and Omori 2003) that by E19–21 the neuronal input resistance, membrane time constant, and action potential half-widths of the tonic neurons are statistically indistinguishable from hatching chicks [postnatal days (P) 3–5]. We argue that our in vitro data, even if slightly more immature, are suitable for answering the questions in the present study.

Whole-cell patch-clamp recordings were made from visually identified NA cells using infrared/differential interference contrast video microscopy. Initial micropipette resistances were 4–8 MΩ with an
In this paper, we estimated the properties of the filter estimated via STA (Recio-Spinoso et al. 2005) and the second-order Wiener kernel exhibits similar properties as the filter estimated via STA (Recio-Spinoso et al. 2005) and can be used to predict responses to sounds (Yamada and Lewis 1999), the second-order Wiener kernel (Schwartz et al. 2006). The STC was derived as the covariance matrix of the spike triggered ensemble. A second-order Wiener kernel (Schwartz et al. 2006). The STC was derived as the covariance matrix of the spike triggered ensemble. C was decomposed in eigenvectors using singular value decomposition. The vector with the second largest singular value, which corresponds to the second-order Wiener kernel, was our estimate of \( h(t) \) for NA neurons. Based on visual inspection, we set the analysis window to be 15 ms, i.e., we considered that the impulse responses \( r(t) \) are shorter than 15 ms, both in ANF and NA. To de-noise the resulting impulse responses, we fitted a linear gamma-chirp to the data (see Fischer et al. 2011; Fontaine et al. 2013 for more details).

Gaussian white noise-based current stimuli were produced using custom routines written in IGOR. A random, Gaussian white noise waveform was convolved with an exponential function \((\tau = 3 \text{ ms})\) and added to a series of constant current steps. Noise amplitude and direct current (DC) level were varied independently. The noise stimulus simulates the arrival of many small, stochastic, and statistically independent synaptic currents, both excitatory and inhibitory, a simplified version of the Ornstein-Uhlenbeck process (Rauch et al. 2003). The time constant of the exponential filter reflects the rapidity of synaptic current events in the auditory brain stem (MacLeod and Carr 2005). These stimuli resemble intracellular membrane potentials \((V_m)\) recorded in cochlear nucleus neurons in vivo (unpublished observations in chicken, K. M. MacLeod, C. E. Carr, and D. Soares; see also (Rhode and Smith 1986)].

Reliability. To compute the reliability of ANF and NA neurons at the time scale of the envelope, we first computed the shuffled autocorrelogram (SAC) (Joris 2003; Louage et al. 2005), the average cross-correlogram between different trials of the same noise. The SACs were normalized by the time bin and the trial duration. Two examples are shown in Fig. 1, C and E. The reliability was defined as in Brette (2012). After removing the baseline of the SAC (equal to \( r^2 \), where \( r \) is the firing rate), the reliability was defined as the normalized integral of the peak:

\[
\text{reliability} = \frac{\int_{-3 \text{ms}}^{3 \text{ms}} \text{SAC}(\tau - r^2) d\tau}{r}
\]

where the integral was taken in a time windows of 3 ms around zero. The reliability is a number between 0 and 1, where 0 is obtained for independent spike trains and 1 when comparing a spike train with a jittered copy (i.e., perfect synchrony if the jitter is 0 ms).

**Reverse correlation on fine structure.** The first element of our model (Fig. 2B, see also Fig. 7A), which was also used to compute the cochlear-induced envelope (Fig. 2C), was a filter that summarizes the linear filtering properties of the afferents to the neuron under investigation. It was characterized by an impulse response, \( h(t) \), which was here calculated using reverse correlation techniques (Schwartz et al. 2006) on spike responses to unfrozen white noise. If the neuron to characterize was sensitive to the sound fine structure, which is the case for barn owl ANF (Köppl 1997), the filter \( h(t) \) was the average stimulus \( \mu \) that elicits spikes, the spike-triggered average (STA) (Boer and Jongh 1978; Carney and Yin 1988; Fischer et al. 2011; Fontaine et al. 2013), that is, \( \mu = 1/N \sum_{n=1}^{N} s(t_n) \), where \( t_n \) is the time of the \( n \)th spike, \( s(t_n) \) is a vector containing the broadband sound stimulus present in temporal window preceding that spike, and \( N \) is the total number of spikes in the analysis. Because most neurons in NA lock very poorly to the sound fine structure, STA could not be used to estimate the filter \( h(t) \). Nevertheless, it was still possible to estimate \( h(t) \). Indeed, as it was shown for neurons that do not lock to sound fine structure (Lewis et al. 2002; Recio-Spinoso et al. 2005; Yamada and Lewis 1999), the second-order Wiener kernel exhibits similar properties as the filter estimated via STA (Recio-Spinoso et al. 2005) and can be used to predict responses to sounds (Yamada and Lewis 1999).

In this paper, we estimated \( h(t) \) using the second filter retrieved via spike-triggered covariance (STC), which is equivalent to the second
EMERGENCE OF BAND-PASS FILTERING THROUGH ADAPTIVE SPIKING

433

stimulus at frequency $f$ that can be accurately reproduced using an optimal linear encoding model. Coherence is a lower bound for information transfer in case of Gaussian signals (Chacron et al. 2004, 2005, 2007; Rieke et al. 1999).

In this paper, we used two kinds of input to compute coherence: the fine-structure signal, $S(t)$, output by the linear filter estimated via reverse correlation, yielding $C_{SR}(f)$ (dashed lines in Fig. 2D) and its envelope $E(t)$, yielding $C_{ER}(f)$ (solid lines in Fig. 2D). To compare the amplitude of the coherences for the two signals, i.e., the degree of information transfer, we took the maximum of $C_{SR}(f)$ around the neuron’s BF for the fine structure signal and at frequencies below 700 Hz for the envelope.

Modulation tuning function. To estimate the modulation tuning function (MTF), we first computed the modulation temporal receptive field (MTRF), which is a linear filter relating the envelope of the input channel to the neuron’s response (Fig. 3A). Our MTRF estimation was based on a recent technique called generalized linear model (GLM) (Calabrese et al. 2011; Paninski 2004; Steinberg et al. 2013) that allows the estimation of receptive fields when the input statistics (in this case the envelope at the output of the cochlear channel) is not white noise. Our GLM implementation is a single-channel monaural

details on the procedure). The best frequency (BF) of a neuron at a given input level is defined as the frequency of the peak of the reverse correlation on fine structure.

Coherence. Coherence analysis (Rieke et al. 1999; Roddey et al. 2000) has been extensively used in the study of envelope coding in the electric fish (Chacron et al. 2007; McGillivray et al. 2012; Middleton et al. 2006). The coherence gives a measure of the filtering properties of a transfer system. Let’s consider an input stimulus $S(t)$ to a neural system, presented $n$ times. The $n$ responses, denoted $R_1(t), R_2(t), ..., R_n(t)$ are spike trains, $R_i(t) = \sum \delta(t - t_i)$. The stimulus-response coherence is

$$C_{SR}(f) = \frac{1}{n} \sum_{i=1}^{n} P_{SR}(f)$$

where $P_{SR}(f)$ is the cross spectrum between $S(t)$ and $R_i(t)$. $P_{SR}(f)$ is the power spectrum of $R(t)$ and $S(t)$, respectively. The coherence, which ranges between 0 and 1, quantifies the degree to which the signal $S(t)$ and $R(t)$ are correlated (both in phase and amplitude) at frequency $f$. Two examples are given in Fig. 2, D and E. The coherence can also be seen as a measure of the fraction of the

---

**Fig. 2.** Coherence analysis of ANF and NA cells. Illustration of the envelope computation is shown. A sound (A) is filtered with the cochlear filter (B) found through spike-triggered techniques. C: the output of the filter comprises a carrier, the fine structure signal (thin line in C), and its corresponding envelope (thick black line in C). D: coherence vs. frequency plots using both the fine structure (dashed line) and envelope (solid line) for an ANF (top graph) and a NA cell (bottom graph). The maximum coherences are shown in (E) for the fine structure and in (F) for the envelope. G: ratio between the maximum envelope coherence $C_{ER}$ and the maximum of the fine-structure coherence $C_{SR}$. Each point represents the response of a cell at a certain input level, which is represented by a circle for NA and a triangle for ANF.

---

**Fig. 3.** Modulation transfer function. Illustration of the modulation tuning function (MTF) computation. A: the generalized linear model (GLM) techniques estimates the linear filter [modulation temporal receptive field (MTRF)] that relates the input envelope and the spiking history to the neuron response after a nonlinearity (NL). B: different characteristics are used to quantify to filter properties: the bandwidth (BW) and the center frequency (CF) of the MTF, the degree of direct current (DC) rejection, and the value of the MTRF negative peak (NP) (C). D and E: MTFs for the entire ANF (D) and NA (E) population. Each row represents the MTF of a response at a certain input level. Dark colors represent higher MTF values. The responses were ordered by CF. NA MTFs are mainly band pass, whereas ANF MTFs are almost always low pass (LP). F: MTFs at different input levels of a particular cell. The NP becomes more negative as the input dB increases. G: the corresponding MTFs. In this example, the degree of DC transmission decreases, and the CF shift to the higher frequencies as input level increases. FFT, fast Fourier transform.
version of the binaural GLM used in Steinberg et al. (2013). The response of a neuron is determined by the conditional intensity function \( \lambda [\mathbf{t} | \mathbf{R}(t)] \), where Prob spike in \((t + \Delta, t + 2\Delta) \) for small time bins \( \Delta \). The response of the neuron \( r(t) \) equals one if there is a spike and zero otherwise. The spike history, \( R(t) \), consists of the response at the previous time steps: \( R(t) = [r(t - \Delta), r(t - 2\Delta), \ldots, r(t - K\Delta)] \). The conditional intensity function is modeled such that (spiking depends on \( I \) the MTRF \( k \) applied to input \( s(t) \); and 2) a spike history term \( k_{\text{spike}} \) that reflects refractory processes: \( \lambda [\mathbf{t} | \mathbf{R}(t)] = \exp [k s(t) + k_{\text{spike}} R(t) + k_j] \) (Fig. 3A). \( s(t) \) is the envelope (computed via a Hilbert transform) of the output of the cochlear filter. The MTRF and the spike history kernel are temporal filters applied to the input signal and the spiking history, respectively. The model parameters, the two vectors \( k \) and \( k_{\text{spike}} \), were found by maximizing the log-likelihood of the data:

\[
L(k) = \sum_{\text{spikes}} \log \left\{ \lambda [t_{\text{spike}} | R(t)] \right\} = -\sum_{t} \lambda [t | R(t)] dt
\]

where \( t_{\text{spike}} \) is spike time. Regularization was used to avoid overfitting by adding a quadratic penalty on the model parameters to the log-likelihood of the data (Steinberg et al. 2015). The regularization parameters for the quadratic penalty were found using cross-validation. Once the MTRF is estimated (Fig. 3B), its Fourier transform gives the MTF (Fig. 3C).

One of the reasons to use a single cochlear channel was that we wanted to compare the MTRFs of the data with those of our models, which only received inputs from one single channel. Because the cochlear filters used to filter the sound are derived from the measurements themselves (by reverse correlation), they reflect better the real inputs than an arbitrary gammatone filter. Still, for control and comparison with previous studies, we also computed MTF from the spectro-temporal receptive fields (STRF; example in Fig. 3). In this case, the STRF was collapsed on its time axis to yield the MTRF. We tried different ways of estimating STRFs: with the normalized STRF technique (Theunissen et al. 2000), with the GLM (Calabrese et al. 2011) and, when the number of spikes was sufficient, using the second-order Wiener kernel (Lewis and van Dijk 2005).

In all cases, the resulting MTFs were qualitatively similar and qualitatively equivalent to our single-channel GLM technique.

Threshold measurement in vitro. Spike threshold was defined as the \( V_m \), at the onset of action potentials (see red points in Fig. 5A), defined as the first time preceding the peak when the first derivative \( dV_m / dt \) crosses a fixed criterion (5 mV/ms) (Azuos and Gray 2000). The threshold is shown as a function of the rate of depolarization over 5 ms preceding a spike (see Fig. 5G) and of the interspike interval (ISI; see Fig. 5H).

STRF separability. To assess the independence of time and frequency tunings, we calculated the separability of the STRFs estimated with GLM (Calabrese et al. 2011). We then computed the separability index (SI) (Depireux et al. 2001; Lesica and Grothe 2008a). First, each STRF was decomposed into its singular values \( k = \sum \lambda_{i,m,v} = \lambda \) are the singular values, and \( u_i \) and \( v_i \) are row vectors corresponding to spectral and temporal cross-sections, respectively, of separable transfer functions. The SI measures the dominance of the first singular value, compared with the other singular values:

\[
SI = \sum_{i} \frac{\lambda_i}{\sum_{i} \lambda_i}
\]

A fully separable STRF has a SI = 1, whereas a fully nonseparable, a SI = 0.

Model. In this paper, the only information we used from neural response was spike timing. As we wanted to investigate how spike generation can modulate the spike responses through threshold adaptation, a detailed subthreshold process model was not needed. We, therefore, used a simple noisy integrate-and-fire to model the subthreshold \( V_m \):
between the model and the data is normalized by each cell’s intrinsic variability between each pair of trials. To take this variability into account, the number of spikes in the experiment and model spike train, respectively.\(\Gamma = \left(\frac{2}{1-2\beta_{exp}}\right)\frac{N_{\text{coinc}} - 2N_{\text{exp}}r_{\text{exp}}}{N_{\text{exp}} + N_{\text{model}}}\)

where \(r_{\text{exp}}\) 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. \(\Gamma = 0\) means that there are no more coincidences than expected by chance, whereas \(\Gamma = 1\) means that the model prediction is perfect, at temporal resolution \(\delta\). We chose \(\delta = 3\) ms, which corresponds to the temporal precision of NA in vitro recordings (Kreeger et al. 2012). The variability of the response of a neuron to the same noise stimulus presented several times can be quantified with the intrinsic gamma factor \(\Gamma_{\text{intra}}\), the mean of all gamma factors computed between each pair of trials. To take this variability into account, \(\Gamma\) between the model and the data is normalized by each cell’s intrinsic gamma factor \(\Gamma_{\text{rel}}\), yielding the relative gamma factor \(\Gamma_{\text{rel}}\).

Because in vivo responses are more noisy than in vitro ones, we did not use the gamma factor, but the coefficient of correlation \(CC\) between the poststimulus timing histogram (PSTH) of the model and the PSTH of the recorded in vivo responses. Also, to capture as much response features as possible, we also used the explained variance \(EV\) of the SAC computed in a 3-ms time windows around zero. In both the in vitro and in vivo case, a term containing the difference between measured and modeled firing rate \(FR\) was added to ease convergence of the optimization algorithm (Fontaine et al. 2013). The final fitness used for optimization is therefore:

\[
\text{fitness(model, data)} = \Gamma(\text{model, data}) + \left|\frac{\text{FR}_{\text{model}} - \text{FR}_{\text{data}}}{\text{FR}_{\text{data}}}ight|
\]

for the in vivo dataset and:

\[
\text{fitness(model, data)} = CC(\text{PSTH}_{\text{model}}, \text{PSTH}_{\text{data}}) + EV(\text{SAC}_{\text{model}}, \text{SAC}_{\text{data}}) + \left|\frac{\text{FR}_{\text{model}} - \text{FR}_{\text{data}}}{\text{FR}_{\text{data}}}ight|
\]

for the in vivo dataset. The optimization used an evolution algorithm called CMAES (Hansen and Ostermeier 2001). The implementation on graphical processing units is described in Rossant et al. (2011). All of the neuron simulations and optimizations were performed using the Brian simulator (Goodman and Brette 2009), the Brian Hears toolbox for auditory filtering (Fontaine et al. 2011), and the Playdoh optimization toolbox (Rossant et al. 2013). All simulations were performed at the same sampling frequency as the recordings, 20 kHz.

**RESULTS**

Our in vivo dataset consisted of single-unit extracellular recordings of 56 NA neurons collected in 4 barn owls and of 126 ANF collected in 6 owls (Fig. 1A). All cells were presented with a set of unfrozen noises to perform reverse correlation, as well as repetitions of the same frozen noise. NA cells were distinguished from NM and ANF units based on their poor ability to phase-lock and their more reproducible spike patterns (Steinberg and Peña 2011). The analyses performed in this paper required thousands of spikes, which was not a problem for ANF, as they tended to fire tonically in response to ongoing stimulus. Nevertheless, because 10% of NA cells are onset (Köppl and Carr 2003), our NA dataset, which only contained tonic cells, was skewed to the most common response type (Köppl and Carr 2003; Steinberg and Peña 2011). Both ANF and NA population spanned the entire tonotopic axis, (800 Hz, 9 kHz) for ANF and (800 Hz, 8 kHz) for NA. Our in vitro dataset consisted of 23 whole-cell patch-clamp recordings in current clamp mode from NA neurons in chick brain stem slices (Kreeger et al. 2012). The stimuli consisted of DC injections of a Gaussian white noise stimuli superimposed on steps of varying mean amplitudes.

**Correlation analysis in ANF and NA responses in vivo.** We compared single-unit extracellular responses of ANF and NA neurons to repeated presentations of a frozen broadband sound. A typical ANF response was weakly reproducible across trials (Fig. 1B), but was strongly time locked to the fine structure of the sound. Evidence for time-locking can be seen in Fig. 1C, where the response average autocorrelogram (calculated as SAC, see MATERIALS AND METHODS) oscillates at the cell’s BF. On the other hand, a typical NA cell displayed a reproducible firing pattern (Fig. 1D), but did not exhibit high-frequency oscillation in its SAC (Fig. 1E), meaning that they lost most of their fine structure encoding capacities (Köppl and Carr 2003; Louage et al. 2005; Steinberg and Peña 2011). It has been suggested that NA reproducible spiking pattern encodes the envelope of the sound (Steinberg and Peña 2011). The reliability of the firing pattern at the time scale of the envelope (3 ms) is shown in Fig. 1F for the entire ANF and NA populations at all input levels. NA displayed higher reliability than ANF (mean ± SD: 0.16 ± 0.11 and 0.05 ± 0.04, respectively). The reliability was correlated with BF in ANF (\(R^2 = 0.42, P < 1\times10^{-10}\)) but not in NA (\(R^2 = 0.007, P = 0.21\)).

**Coherence analysis in ANF and NA neurons.** Even if the sound presented does not have an envelope (Fig. 2A), the signal transmitted to the brain does (Joris 2003; Louage et al. 2005). The cochlea acts as a band-pass filter (Fig. 2B), and, when a noise is passed through a band-pass filter, the output \(S(t)\) can be described as the product of a carrier with the same frequency as the neuron’s BF (thin line in Fig. 2C) and a modulating signal \(E(t)\) (thick line in Fig. 2C). \(S(t) = E(t)\cdot\sin(2\pi BF\cdot t + \theta)\), where BF is the that of the cell, and \(\theta\) is a random phase. In this study, “envelope” refers to this cochlea-induced envelope \(E(t)\), while “fine structure signal” to \(S(t)\). For each neuron, we estimated the cochlear filter (Fig. 2B) using the STA for ANFs and using covariance analysis for NA (even if NA cells do not phase lock, it is still possible to estimate the fine structure filter using covariance analysis (see MATERIALS AND METHODS; Lewis et al. 2002; Recio-Spinoso et al. 2005; Yamada and Lewis 1999)).

The information about the stimulus contained in the response of a single cell can be estimated using coherence analysis (Chacron et al. 2005; McGillivray et al. 2012; Middleton et al. 2006). The coherence (see MATERIALS AND METHODS for more details), which ranges between 0 and 1, quantifies the degree to which two signals are linearly correlated (both in phase and amplitude) at frequency \(f\). The coherence can also be seen as a measure of the fraction of the stimulus at frequency \(f\) that can be accurately reproduced using an optimal linear encoding model. We used this technique using both the envelope \(E(t)\) and the fine structure signal \(S(t)\) as input, and the spike train responses to frozen noise as output. Examples are given in Fig. 2D for a single ANF (top graph) and for a single NA cell (bottom graph) with the coherence between fine structure and response (\(C_{SR}\) dashed line) and the coherence between envelope and response (\(C_{ER}\) solid line). The amount of information was quantified by taking the maximum of the coherence function below 800 Hz for the envelope (max \(C_{SR}\) in Fig. 2D) and around the cell’s BF for the fine structure signal (max \(C_{ER}\) in Fig. 2D). Those measurements for all the responses at all input
levels are plotted in Fig. 2E for the fine structure signal for both ANF (triangles) and NA (circles) and for the envelope in Fig. 2F. As previously shown in barn owls (Köppl 1997) and also in mammals [e.g., in cat (Louage et al. 2005)], the ability of ANFs to encode the fine structure of a sound degraded as BF increased. This behavior was the same for the envelope. Moreover, ANFs had higher coherence values for the fine structure than for the envelope for BF < 4 kHz, whereas the coherence is similar for higher frequency. The ability of NA to encode fine structure, quantified by the coherence, was insignificant for BF > 2.5 kHz and very low compared to ANFs < 2.5 kHz (Köppl and Carr 2003). Nevertheless, the envelope coherence of NA was high, much higher than for ANFs. To further illustrate the difference between ANF and NA coherences, the ratio between $C_{ER}$ and $C_{ER}$ is plotted in Fig. 2G. Those results demonstrate that, at all BFs, a single NA cell contained more information about the envelope than an ANF.

**Modulation transfer function.** We used both time and frequency representations to characterize the filtering performed on the envelope (Lesica and Grothe 2008b; Rodriguez et al. 2010a, 2010b; Woolley et al. 2005). The time representation, called the MTRF, is the impulse response of the linear filter that relates the input amplitude modulation to the neuron’s response (Fig. 3A). The corresponding frequency representation, called the MTF, reflects the neuron’s response to different modulation frequencies (Fig. 3C). To quantify the filter properties, in particular whether it was low pass or band pass, we calculated the bandwidth (BW) of the MTF, its center frequency (CF), the value at frequency zero of the normalized MTF that we call DC transmission (DCtrans), which can range from 0 (complete DC rejection) to 1 (pure low pass) (Fig. 3C), as well as the value of the negative peak (NP) of the normalized MTF (Fig. 3B).

As shown in Fig. 3D, where every row represents the MTF of a response, the ANF MTFs were always low pass (26 out of 301 responses have DCtrans < 1 with mean 0.68), whereas the NA MTFs were mainly band pass (111 out of 176 responses have DCtrans < 1, mean 0.30). The NA envelope filters depended on the input level. An example of a cell presented with stimuli at different levels is shown in Fig. 3F for the time representation (MTRFs) and in Fig. 3G for the frequency representation (MTFs). The NP of the MTRF became more negative as the input level increases. The MTFs were more low pass at low input level and became band pass as the level increases. The relationships of the filtering features to sound level are shown in Fig. 4, A–D, for a subpopulation of 13 NA neurons where responses of each single cell at different input levels are joined together by a line. Similar to what has been found in the mammalian inferior colliculus (IC) (Krishna and Semple 2000; Lesica and Grothe 2008a) and the avian field L (Nagel and Doupe 2006), NA MTFs were more low pass at low input levels and became band pass as the level increased [DCtrans decreased (Fig. 4C), and NP became more negative (Fig. 4D)]. We defined the sensitivity of a filtering feature to input level as the slope of the regression line for each individual cell. For the whole NA sample ($n = 56$), CF increased with input level (mean sensitivity $24 \pm 9.3$ Hz/dB; significantly greater than zero, one-tailed $t$-test, $P < 0.0001$), DCtrans decreased with input level (mean sensitivity $-0.15 \pm 0.06$ Hz/dB, $P = 0.0001$), NP became more negative with input level (mean sensitivity $-0.15 \pm 0.03$ Hz/dB, $P = 1\text{e}^{-8}$), and BW increased with input level (mean sensitivity $-27 \pm 11$ Hz/dB, $P = 0.0001$). These results confirmed the change in filter properties as a function of input level.

It was shown that the negative part of the MTRF, hence their filter properties, correlated with the spike pattern reproducibility across frozen noise trials (Steinberg and Peña 2011). In particular, the more negative was the bump of the MTRF, the more reproducible were the responses. We tested here whether the encoding efficiency, measured via coherence analysis, depended on the filtering properties of a cell at a given input level. To do so, we calculated for each cell the sensitivity of the maximum envelope coherence maximum $C_{ER}$ to the filtering features, as shown for DCtrans in Fig. 4E. For the whole sample are shown in Fig. 4F for NP and Fig. 4G for DCtrans. The coherence increased as NP became more negative (mean sensitivity $-0.70 \pm 1.7$, $P = 0.008$). The coherence increased as DCtrans decreased (mean sensitivity $-0.41 \pm 0.6$, $P = 0.0001$). The coherence did not significantly depend on BW or CF. These results demonstrated that the information transfer about the envelope increased when the envelope was band-pass filtered, which occurred when input level increased.

NA neurons show band-pass filtering in vitro. To explore whether the transformation of MTF properties from low-pass
in ANFs to band-pass in NA originates in cellular mechanisms, we used an in vitro approach. NA neurons were recorded in current clamp using current injections that mimicked the ANF input at different sound levels (Fig. 5A). In this paradigm, mean injected-current amplitude corresponded to sound input level and current variance to input synchrony from ANFs (Kreeger et al. 2012). The cells of interest for the present study were those previously defined as differentiator in Kreeger et al. (2012) (n = 16), where firing rate increased with input-current variance, even at high mean input levels. We analyzed the filter properties of each neuron similarly to our in vivo analysis (Fig. 5) using a standard spike-triggered average. The effect of changing input intensity was similar to in vivo recordings. NP (Fig. 5B) became more negative, and DCtrans (Fig. 5C) decreased with intensity. Across the entire range of stimulus means, 66 MTFs of 86 responses showed band-pass MTFs, with DCtrans = 0.44 ± 0.25 (Fig. 5C). Representative plots of DCtrans vs. input level are shown for five NA neurons in Fig. 5D. The effect of input level on the MTF properties was consistent for the whole sample (n = 16). DCtrans decreased with mean input current (mean sensitivity of −1.59 ± 1.74 Hz/nA, P < 0.001). The NP of the STA became more negative when the mean input current increased (mean sensitivity of −1.48 ± 1.13 Hz/nA, P < 0.001). CF increased (91 ± 76 Hz/nA, P = 0.002), while BW did not change significantly with mean input current (mean sensitivity of 16 ± 46 Hz/nA, P = 0.17). These results were consistent with the changes of filter properties as a function of input level measured in vivo and, because they relied only on single-cell mechanisms, suggested that neither network [e.g., delayed inhibition (Nelson and Carney 2004)] nor synaptic (Fortune and Rose 2000) mechanisms were necessary to elicit the MTF behavior observed in vivo.

Subthreshold resonance has also been suggested as a candidate mechanism to implement band-pass filtering (Hutcheon and Yarom 2000) such that ionic-conductance dynamics influence the MTF properties. As the speed of depolarization preceding a spike (Fig. 5G) increased (91 ± 76 Hz/nA, P = 0.002), while BW did not change significantly with mean input current (mean sensitivity of 16 ± 46 Hz/nA, P = 0.17). These results were consistent with the changes of filter properties as a function of input level measured in vivo and, because they relied only on single-cell mechanisms, suggested that neither network [e.g., delayed inhibition (Nelson and Carney 2004)] nor synaptic (Fortune and Rose 2000) mechanisms were necessary to elicit the MTF behavior observed in vivo.

Subthreshold resonance has also been suggested as a candidate mechanism to implement band-pass filtering (Hutcheon and Yarom 2000) such that ionic-conductance dynamics influence the MTF properties. As the speed of depolarization preceding a spike (Fig. 5G) increased (91 ± 76 Hz/nA, P = 0.002), while BW did not change significantly with mean input current (mean sensitivity of 16 ± 46 Hz/nA, P = 0.17). These results were consistent with the changes of filter properties as a function of input level measured in vivo and, because they relied only on single-cell mechanisms, suggested that neither network [e.g., delayed inhibition (Nelson and Carney 2004)] nor synaptic (Fortune and Rose 2000) mechanisms were necessary to elicit the MTF behavior observed in vivo.

Subthreshold resonance has also been suggested as a candidate mechanism to implement band-pass filtering (Hutcheon and Yarom 2000) such that ionic-conductance dynamics influence the MTF properties. As the speed of depolarization preceding a spike (Fig. 5G) increased (91 ± 76 Hz/nA, P = 0.002), while BW did not change significantly with mean input current (mean sensitivity of 16 ± 46 Hz/nA, P = 0.17). These results were consistent with the changes of filter properties as a function of input level measured in vivo and, because they relied only on single-cell mechanisms, suggested that neither network [e.g., delayed inhibition (Nelson and Carney 2004)] nor synaptic (Fortune and Rose 2000) mechanisms were necessary to elicit the MTF behavior observed in vivo.

Subthreshold resonance has also been suggested as a candidate mechanism to implement band-pass filtering (Hutcheon and Yarom 2000) such that ionic-conductance dynamics influence the MTF properties. As the speed of depolarization preceding a spike (Fig. 5G) increased (91 ± 76 Hz/nA, P = 0.002), while BW did not change significantly with mean input current (mean sensitivity of 16 ± 46 Hz/nA, P = 0.17). These results were consistent with the changes of filter properties as a function of input level measured in vivo and, because they relied only on single-cell mechanisms, suggested that neither network [e.g., delayed inhibition (Nelson and Carney 2004)] nor synaptic (Fortune and Rose 2000) mechanisms were necessary to elicit the MTF behavior observed in vivo.

Threshold adaptation can explain band-pass filtering. It has been suggested that the differentiator properties of NA (Kreeger et al. 2012) and pyramidal cells (Arsiero et al. 2007) can originate in threshold adaptation, which has been linked to high-pass filtering (Benda et al. 2010; Higgs and Spain 2011; Platkiewicz and Brette 2011). To determine whether spike threshold adaptation occurred in NA neurons, we directly measured thresholds from the voltage responses in NA neurons in vitro (red dots in Fig. 5A; see MATERIALS AND METHODS). Threshold was quite variable; in the histogram for one neuron shown in Fig. 5F the threshold had a variance of 7 mV. As reported in other central, e.g., in visual (Arousey and Gray 2000), auditory (Péna and Konishi 2002), and somatosensory (Wilent and Contreras 2005) systems, the threshold was correlated to the rate of depolarization preceding a spike (Fig. 5G), suggesting it adapts to $V_{m}$ (Platkiewicz and Brette 2010). Also, as reported for instance in the hippocampus (Henne and Buz-
sáki 2001), the threshold was inversely correlated with the ISI (Fig. 5H), suggesting that the threshold increased after each spike and relaxed back with a time constant that was longer than the typical ISI (Benda et al. 2010; Chacron et al. 2007). To test whether threshold dynamics correlated with MTF properties, we calculated the sensitivity of the MTFs features to threshold variability (TV), defined as the ratio between the variance of spike onsets and the variance of $V_m$ for each input mean (TV = 0.2 ± 0.16 for the population). The DCTrans of the MTF decreased with increased TV (example in Fig. 5I). Over all neurons, both DCTrans and NP decreased with TV (mean sensitivity $-3.2 \pm 4.1$, $t = -2.4$, $P = 0.02$, and $-3.77 \pm 4.84$, $t = -2.3$, $P = 0.02$, respectively), showing that the filtering became more band pass as the threshold increased in variability.

How could threshold adaptation change neural filtering properties? The case of spike-triggered threshold adaptation was explained in details in Benda et al. (2010). Briefly, if stimulus features are faster than the adaptation process (high-frequency components), the neuron’s threshold does not have time to adapt, and spikes can more easily be triggered by them. If the stimulus components are slower than the adaptation time constant (low-frequency components), the neuron reaches a fully adapted state, and spikes are less likely to be elicited. Because, for a given input current amplitude, a neuron fires less in the adapted state, the output firing will be higher for high frequencies, thus leading to high-pass filtering. Coupled with the intrinsic low-pass filter properties of the membrane, this phenomenon generates band-pass filtering. To explain the effects of the threshold adaptation to the $V_m$, let us consider an adaptive process where the threshold $V_t(t)$ depends on the $V_m(t)$ through a first order differential equation (Farries et al. 2010; Fontaine et al. 2014; Higgs and Spain 2011; Platkiewicz and Brette 2010):

$$\frac{dV_t(t)}{dt} = -V_t(t) + \theta_s(V_m)$$

where $\tau_t$ is the threshold time constant, and $\theta_s(V_m)$ is the threshold steady-state function (taken as the identity for the current example). The threshold is therefore a low-pass version of $V_m$. If the threshold adapts, the effective signal in triggering a spike is no longer $V_m(t)$, but the distance between $V_m(t)$ and $V_t(t)$. This corresponds to a neuron with a fixed threshold (FT) for which the subthreshold signal is $V_m(t) - V_t(t)$ (Fig. 6A). Subtracting from a signal a low-pass version of itself is the same as high-passing it (Fig. 6B). The filter chain before spike generation is shown in Fig. 6C: the signal is low-pass filtered by the synapses (blue), then by the membrane (green), then high-pass by the threshold (red). The overall filtering effect is therefore band pass (dash purple line in Fig. 6C).

It has been proposed that TV reflects adaptation of the spike threshold to the $V_m$ due to the inactivation of sodium channels (Hu et al. 2009; Kuba et al. 2010; Platkiewicz and Brette 2011).

Fig. 6. Adaptive threshold model. A: if the spike threshold varies in time, the effective signal in triggering spikes is the distance between the $V_m$ and the threshold. This corresponds to a neuron with this distance as subthreshold signal and with a fixed spike threshold. B: because the threshold is a LP version of the $V_m$, the threshold actually acts as a high pass (HP) in the filter chain. C: the total filtering resulting from the synapses, the membrane (both LP), and the threshold (HP) is band pass. D and E: this model (without synapse) is fitted to a cell responding to two currents with different means [low intensity (D) and high intensity (E)]. D1 and E1: the resulting $V_m$ values (in blue) and thresholds (in green). The red points are the measured spikes. D2 and E2: the resulting modeled STAs in red compared with the measured one (in blue). D3 and E3: the threshold ($V_t$) is more variable for the high intensity because the corresponding $V_m$ distribution falls in a region where the threshold steady-state function ($\theta_s$) varies (E3) unlike for the low intensity (D3).
to the activation of potassium channels (Goldberg et al. 2008; Higgs and Spain 2011). As in our data, it was shown that hyperpolarizing a cell in vitro removes the negative component on the STA (therefore passing from a band pass to a low pass) in another auditory brain stem nucleus (Svirskis 2004). This effect was attributed to sodium inactivation, as it was still present when the potassium currents were blocked. A model of threshold adaptation based on sodium channel kinetics previously developed (Platkiewicz and Brette 2011) was, therefore, used here to test whether threshold adaptation can explain the in vitro and in vivo MTF properties.

The spiking neuron model was a noisy leaky-integrate and fire neuron with an adaptive threshold that depended on the subthreshold \( V_m \) (Farries et al. 2010; Fontaine et al. 2014; Higgs and Spain 2011; Platkiewicz and Brette 2010) and on the firing history (Benda et al. 2010; Chacron et al. 2007) (see MATERIALS AND METHODS). The threshold steady state function \( \theta(V_m) \) converges to a fixed value at low \( V_m \) (Higgs and Spain 2011; Platkiewicz and Brette 2010) and, therefore, can be simplified by a rectified piecewise linear function (Fig. 6, D3 and E3) (Platkiewicz and Brette 2011). Below a given voltage \( V_r \), all Na channels are available, and the threshold remains constant. As the voltage increases, more and more Na channels deactivate, and the threshold increases. We fitted the model to a cell recorded in vitro at two mean input levels (Fig. 6, D and E, see MATERIALS AND METHODS). In Fig. 6, D1 and E1, the resulting \( V_m \) values (in blue) and dynamic thresholds (in red) are shown. The red dots indicate firing times of the measurements. We observed an almost perfect fitting of the model to the data (relative gamma factor of 0.93). The corresponding MTFs were shown in Fig. 6, D2 and E2, where there was a very good correspondence between the model and the measurements. The shape of the predicted threshold steady-state functions \( \theta(V_m) \) shown in Fig. 6, D3 and E3, was essential to explain the data. Indeed, at low mean input intensity, the distribution of the subthreshold membrane voltage was within a range where \( \theta(V_m) \) is constant (Fig. 6, D3) and threshold adaptation to the \( V_m \) was low (Fig. 6, D1). Moreover, spike-triggered threshold adaptation was also low because the firing rate is low. For the higher mean input level, the \( V_m \) distribution was within a range where \( \theta(V_m) \) varies (Fig. 6, E3), yielding more TV to \( V_m \) (Fig. 6, E1). As the firing rate was higher, spike-triggered threshold adaptation was also higher. The linear filtering properties of threshold adaptation presented in Fig. 6C could, therefore, become nonlinearly dependent on mean \( V_m \), through a nonlinear \( \theta(V_m) \). Note that the model could fit very reliably the whole population of cells recorded in vitro (relative gamma factor of 78%, see MATERIALS AND METHODS).

**In vivo response modeling.** While in the literature, band-pass MTFs have often been modeled as delayed inhibition (Nelson and Carney 2004; Rodriguez et al. 2010b), our in vitro and theoretical analysis demonstrated that a single-cell mechanism consisting of an adaptive threshold could explain these features. We, therefore, went back to the in vivo recordings to test whether the responses and the corresponding MTFs could be predicted by the model.

The cellular model proposed in this paper assumed that the preferred acoustic frequency of NA (along its tonotopical axis) has presynaptic origin, it depends on the frequency tuning of ANF inputs, and that the modulation filtering, i.e., the MTF properties, was due to NA cellular mechanisms. In other words, the two types of filtering, frequency and modulation selectivity, should be independent of each other. If this were the case, the STRFs (one of which is shown in the inset of Fig. 3B) should be fully separable, i.e., they could be inferred based on a pair of temporal and spectral transfer functions (Depireux et al. 2001). Separability can be measured using the SI (see MATERIALS AND METHODS), which ranges from 0 for nonseparable STRF to 1 for fully separable ones. For our NA population, the SI was very high (mean SI = 0.98 ± 0.02). It was shown in the mammalian IC (Lesica and Grothe 2008a) that STRFs are highly separable at low input levels (SI around 0.95) but decrease at high input levels (SI around 0.7). NA is different as the STRFs are separable at all input levels.

The preprocessing chain that generated the input to the NA cell model is shown in Fig. 7A. The sound passed through the cochlear filter, estimated from NA responses to unfrozen noise (see MATERIALS AND METHODS). The output was half-waved rectified, which yields a good approximation of ANF PSTHs (Boer and Jongh 1978; Carney and Yin 1988). The half-wave rectification performed by the hair cells is modeled as a nonlinear operator that extracts the envelope from the signal.
(Savard et al. 2011), i.e., the output of the half-wave rectifier contains energy in its spectrum corresponding to the envelope components that were not present at the output of the filter. To model the decrease in the vector strength of phase-locking of ANF (Köppl 1997), rectified signal was low-pass filtered with a cut-off frequency of 3.5 kHz, estimated from the ANF dataset. At this stage in the model, corresponding to the output of the ANFs, the envelope is extracted through half-wave rectification, and the higher frequency components are already attenuated due to the low-pass properties of the hair cells.

Because each NA cell receives input from many ANFs (MacLeod and Carr 2005), we assumed that the output of the ANF stage approximates the PSTH of all pooled ANF inputs to the NA. This signal was scaled, and an offset was added to mimic realistic NA inputs (Kreeger et al. 2012). Indeed, the sound-driven potentials appear as rapid fluctuations superimposed on a DC pedestal (unpublished observations, K. M. MacLeod, C. E. Carr, D. Soares), similar to those reported from mammalian cochlear nucleus (Rhode and Smith 1986). Because the mean input in vivo depends on sound level, we assumed that the output of the half-wave rectification, and the higher frequency components are already attenuated due to the low-pass properties of the hair cells.

We restricted our modeling to cells that met the criteria of being band pass with sound level and displayed monotonic rate-level functions. Also, the cells needed to exhibit the typical NA reproducible firing pattern. We, therefore, chose cells with SACs yielding coincidence coefficients larger than 3 (Louage et al. 2005; Steinberg and Peña 2011; see MATERIALS AND METHODS). This procedure yielded $n = 38$ cells out of 56, which corresponded to the fraction of tonic differentiator cells found in NA in vitro (Kreeger et al. 2012). The model fitting (see MATERIALS AND METHODS for more details) was performed using responses at two sound levels. For each cell, the level difference for the two responses was the largest possible (with a minimum of 10 dB). The first 50 ms were discarded to avoid transient artifacts. The next 200 ms were used for model learning, and the last 250 ms for testing. An example cell stimulated at 55 dB and 75 dB is shown in Fig. 7, C–G. The resulting membrane voltage and threshold are shown in Fig. 7C for each level. The threshold was less variable at the low input level. The steady-state function (Fig. 7D) had, in the working range of $V_{\text{m}}$, a constant and a variable region. The SACs of both modeled and measured responses were similar at both input levels (Fig. 7E). A small oscillation in the modeled SAC at low level indicates that the fine structure encoding did not completely disappear in the model. The measured response (Fig. 7F top row) was very similar to the predicted one (Fig. 7F, bottom row; CC between measured and predicted PSTHs is 0.76). More importantly, the MTF, which was more band pass at high input level, was very well predicted by the model (Fig. 7G).

The same model fitting was performed on the entire population. The resulting parameter statistics are given in Table 1. The mean CC of measured vs. predicted PSTH was $0.69 \pm 0.12$ for the whole sample (compared with $0.5 \pm 0.17$ for a model with FT) and was not correlated with the cells’ BFs ($R^2 = 0.001$, $P = 0.81$). The response properties of the predicted response are plotted against the corresponding measured ones in Fig. 8, A–F, where each point represents a cell’s response at a given input level. The model predicted well the features of interest: reliability ($R^2 = 0.63$, $P < 1e-8$ compared with $R^2 = 0.22$, $P = 0.009$ with FT) (Fig. 8A), maximum envelope coherence ($R^2 = 0.68$, $P < 1e-12$ compared with $R^2 = 0.39$, $P < 1e-3$ with FT) (Fig. 8B), filter CF ($R^2 = 0.68$, $P < 1e-12$ compared with $R^2 = 0.65$, $P < 1e-3$ with FT) (Fig. 8C), degree of DC transmission ($R^2 = 0.67$, $P < 1e-12$ compared with $R^2 = 0.1$, $P = 0.51$ with FT) (Fig. 8D), NPs of MTRF ($R^2 = 0.48$, $P < 1e-8$ compared with $R^2 = 0.05$, $P = 0.05$ with FT) (Fig. 8E), filter BW ($R^2 = 0.68$, $P < 1e-12$ compared with

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Means ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{\text{refr}}$, ms</td>
<td>5 ± 2.5</td>
</tr>
<tr>
<td>$R_{\text{C}}$, MΩ</td>
<td>416 ± 134</td>
</tr>
<tr>
<td>$V_{\text{r}}$, mV</td>
<td>$-71 \pm 7$</td>
</tr>
<tr>
<td>$\sigma$, mV</td>
<td>9.1 ± 1.3</td>
</tr>
<tr>
<td>Refractory, ms</td>
<td>3.3 ± 0.8</td>
</tr>
<tr>
<td>$\Delta$, ms</td>
<td>0.4 ± 0.3</td>
</tr>
<tr>
<td>$\alpha$, mV</td>
<td>4.9 ± 3</td>
</tr>
<tr>
<td>$\tau_{\text{syn}}$, ms</td>
<td>17 ± 12</td>
</tr>
<tr>
<td>$\sigma_c$, mV</td>
<td>0.64 ± 0.23</td>
</tr>
<tr>
<td>$V_{\text{c}}$, mV</td>
<td>$-71 \pm 3$</td>
</tr>
<tr>
<td>$V_{\text{r}}$, mV</td>
<td>$-66 \pm 9$</td>
</tr>
<tr>
<td>$\tau_{\text{DW}}$, μs</td>
<td>530 ± 120</td>
</tr>
<tr>
<td>$S_{\text{Cf}}, \text{pA/} \text{dB}$</td>
<td>9 ± 6</td>
</tr>
<tr>
<td>$S_{\text{Cf}, \text{NP}}, \text{pA/} \text{dB}$</td>
<td>92 ± 5</td>
</tr>
</tbody>
</table>

See text for definition of terms.
\[ R^2 = 0.55, P < 1e-3 \text{ with FT} \] (Fig. 8F). These results showed that the explained variance of most of the filter features was higher than 60%. Although the FT model could predict the PSTHs quite reliably, its prediction performance of the MTF properties and coherence values was very low, demonstrating that the dynamic processes underlying adaptive threshold were necessary to shape neural filtering and information transfer.

In light of the in vitro observations, we wanted to know whether the threshold dynamics were related to the MTF properties. To do so, we checked whether the degree of threshold variation correlated with the MTF features in each cell. Unlike for the in vitro recordings where threshold could be only estimated at spike onset, our model yielded a threshold value at all times. This continuous value was used to compute the TV. The results are shown in Fig. 9, where it can be seen that the MTF features of the recorded responses were indeed correlated with the TV: filter CF (\( R^2 = 0.43, P < 1e-12 \)), regression line \( 190x + 58 \) Hz (Fig. 9A), degree of DCtrans (\( R^2 = 0.51, P < 1e-10 \)), regression line \(-0.44x + 0.6\) Hz (Fig. 9B), NP of MTRF (\( R^2 = 0.53, P < 1e-12 \)), regression line \(-0.37x - 0.11\) Hz (Fig. 9C), filter BW (\( R^2 = 0.41, P < 1e-12 \)), regression line \(167x + 167\) Hz (Fig. 9D). These results indicated that the dynamic processes responsible for TV determine the MTF properties.

**Adaptation time scale.** A fundamental feature in sensory systems is neural adaptation to stimulus statistics (Brenner et al. 2000; Fairhall et al. 2001; Lesica and Grothe 2008a; Nagel and Doupe 2006). In the auditory system, it has been shown that response properties, e.g., STRFs, change when the sound level (its mean) or contrast (variance) changes abruptly in a continuous stimulus (Lesica and Grothe 2008a; Nagel and Doupe 2006). However, in those studies, it was not possible to test whether the response changes occurred on a time scale smaller than 100 ms. Our model predicts that the MTF features should vary with input level on a time scale similar to the threshold time constant, i.e., around 15 ms. To test this prediction, instead of changing the input level abruptly, we used a ramped stimulus that increased progressively from 0 dB to 90 dB in 150 ms and back down to 0 dB in 450 ms (Fig. 10). We then computed MTFs with consecutive partially overlapping 10-ms time windows along the ramp. Because there were not many spikes per trial per time window, the stimulus was repeated at least 2,000 times. One example is shown in Fig. 10B. As predicted, the MTFs properties followed the instantaneous change of input level. In particular, it was low pass at low level and band pass at high level. The correlation coefficients between the instantaneous input levels and the time-varying MTF parameters of eight cells presented with the ramp was significant (DCtrans = 0.52 ± 0.16, BW = 0.4 ± 0.27, NP = 0.66 ± 0.18, CF = 0.66 ± 0.21, \( P < 0.01 \)).

**DISCUSSION**

Envelope encoding is of ethological value in sensory processing (Attias and Schreiner 1997; Mante et al. 2005; McGilivray et al. 2012). In audition, it is essential in speech (Shannon et al. 1998) and sound recognition (Nelson and Takahashi 2010; Schneider and Woolley 2010; Singh and Theunissen 2003). We showed that NA cells encoded the envelope of the cochlear channel outputs more efficiently (Fig. 1F) and reliably (Fig. 2F) than ANFs, consistent with a role of CN in envelope coding enhancement (Frisina 2001; Frisina et al. 1997; Gai and Carney 2008). A single-cell mechanism based on the spike generation nonlinearity could account for the emergence of level-dependent selectivity to modulation frequency, i.e., the filtering became more band pass as the input level increased. Furthermore, we found that threshold adaptation in vitro was consistent with the neural filtering of NA cells (Fig. 5I).

Threshold adaptation to the \( V_m \) has numerous functional implications, such as enhancing input coincidence detection (Azouz and Gray 2000, Fontaine et al. 2014) and improving feature selectivity (Gittelman et al. 2009; Wilent and Contreras 2005), while spike-triggered threshold adaptation shapes information transfer and improves robustness to noise (Chacron et al. 2004, 2007). In the electrosensory system, it was shown that threshold adaptation can extract the envelope (Middleton et al. 2006). While frequencies corresponding to the envelope are already present in the spectrum of the inputs to NA (ANF in

---

**Fig. 9.** Measured MTF properties vs. model threshold variability. The x-axis represents the threshold variability defined as the ratio between the variance of the threshold and the variance of the membrane voltage. The y-axis represents MTF properties of the measured NA responses. A: MTF CF. B: MTF DC. C: MTRF NP. D: MTF BW.

**Fig. 10.** MTFs as a function of instantaneous input level. A: the stimulus is a frozen noise that ramps up from 0 dB to 90 dB in 150 ms and ramps down back to 0 dB in 400 ms. The noise is repeated at least 2,000 times. B: example of MTFs as a function of time for one NA unit. Each column is an MTF as a function of frequency. MTF were computed in 10-ms time windows that overlap by 50%.
Fig. 2), this process could play a role increasing the efficiency of envelope encoding in NA’s output. The magnitude of the coherence, which gives a lower bound for information transfer (Chacron et al. 2005), was higher in NA than in ANF for the envelope but not for the fine structure. Also, for a given cell, coherence increased as filtering became more band pass, an effect predicted in our model by increased threshold adaptation. We also showed that the coherence in modeled responses was higher with an adaptive threshold than with a FT (Chacron et al. 2007). These results suggest that the dynamic processes underlying threshold adaptation can play a role in increasing envelope encoding efficiency between ANF and NA.

Threshold adaptation as a mechanism for band-pass filtering. Spike threshold adaptation in vivo plays an important role shaping MTFs and STRFs in the mammalian IC (Geis and Borst 2009; Gittelman et al. 2009; Gittelman and Pollak 2011). We showed in particular that threshold adaptation can affect the transmission of DC components of the signal. It has been shown in vitro (Higgs and Spain 2011) and theoretically (Platkiewicz and Brette 2011) that threshold adaptation can act as a high-pass filter. We provided an explanation how band-pass filtering can emerge via threshold. Even though we cannot be conclusive about the origin of threshold adaptation, whether sodium (Fontaine et al. 2014; Hu et al. 2009; Kuba et al. 2010) or potassium (Goldberg et al. 2008; Higgs and Spain 2011) channel dynamics underlie the phenomenon, we showed that a model representing sodium channels with appropriate biophysical inactivation properties can explain the data. Indeed, the shape of the threshold steady-state function was crucial in explaining the dependence of the filtering on input level. At low input ranges, all sodium channels are available, and the threshold will be more or less constant, leaving the low-pass membrane as unique filtering component (Fig. 6D). When the level increases, the threshold becomes variable, inducing band-pass filtering (Fig. 6E). Threshold adaptation in our model emerges from ionic mechanisms responsible for spike generation. Because spike generation is always present in firing neurons, this functional adaptation does not add complexity at the cellular or circuit level nor increase energy consumption.

Spike-triggered current adaptation could have a similar effect on the transfer function to that of spike-triggered threshold adaptation (Benda et al. 2010). However, differences between the two processes exist. Spike-triggered current adaptation has a subtractive effect on the frequency-current (f-I) curve, whereas spike-triggered threshold adaptation has a divisive effect (Benda et al. 2010). Therefore, adaptation currents tend to linearize the steady-state f-I curve, whereas dynamic thresholds make them more nonlinear. While not entirely conclusive, the departure from linearity of NA f-I curve (Fig. 2A of Kreeger et al. 2012), combined with evidence from direct threshold measurements (Fig. 5H), suggests that spike-triggered threshold adaptation is prominently present in differentiator NA cells.

Alternative band-pass filtering mechanisms. Because the change of filtering properties with input statistics is very fast, it has been argued that the underlying mechanism should involve a change in static nonlinearity (Lesica and Grothe 2008a; Nagel and Doupe 2006). Such static nonlinearities can emerge through interplay between excitation and inhibition. A neuron receives feed-forward inhibition with the same temporal and spectral tuning as excitation, but with inhibition slightly delayed, inducing a negative bump in the MTRF (Nelson and Carney 2004). If inhibitory afferents have a higher threshold, the MTF will be level dependent. Inhibition has been shown to enhance envelope synchrony (Gai and Carney 2008). Even if a role of inhibition in shaping in vivo MTFs cannot be excluded, our in vitro somatic recordings, which presumably do not reflect network effects, were consistent with the in vivo analysis, rendering the delayed inhibition hypothesis unnecessary to explain our data. The measured changes of MTF were indeed very fast, but, unlike what has been argued (Lesica and Grothe 2008a; Nagel and Doupe 2006), they were consistent with threshold adaptation time constants (Fig. 10B). We thus conclude that, even if they might co-exist with delayed inhibition in NA cells, cellular processes play a role in shaping MTF properties.

The band-pass property of a neural transfer function can emerge from other mechanisms. A synapse capable of both short-term depression and facilitation can become a band-pass filter (Fortune and Rose 2001). Depression acts as a low-pass filter, attenuating the effect of high-frequency presynaptic firing, whereas facilitation acts as a high-pass filter by enhancing the efficacy of presynaptic firing. Short-term synaptic plasticity has been reported at the ANF to NA synapses in vitro, which could affect the filtering properties of these connections (MacLeod et al. 2007; MacLeod and Horiiuch 2011). Whether similar plasticity contributes to level-dependent changes in the modulation transfer function has yet to be determined.

Band-pass filtering can also be a consequence of subthreshold resonance, where a current opposing low-frequency Vm fluctuations acts as a high-pass filter (Hutcheon and Yarom 2000). In the auditory brain stem, low-threshold potassium current (Kth) has the kinetics to induce resonant frequencies as high as the ones measured in NA neurons (Meng et al. 2012; Svirskis 2004). Onset firing in response to a step current, the hallmark of Kth currents (Oertel 1999), was reported in only 17% of NA neurons (Fukui and Ohmori 2003), suggesting that Kth might not be expressed at sufficient levels to suppress firing in a majority of NA cells. Moreover, our results showed that only 15% of the in vitro subthreshold responses exhibited a shallow resonance (Fig. 5E); potentially due to Ksh, whereas 78% of the responses were band pass. Thus resonance due to Kth, while potentially present in some cells, does not seem to be the main mechanism explaining band-pass filtering in NA cells.

Functional implications for sound processing. While different techniques have been used across studies to measure the tuning of neurons to envelope modulation frequency, it is informative to compare our results with previous reports. In mammals, envelope coding changes along the ascending auditory system (Joris 2004). In particular, MTFs become band pass at the level of the IC. Our results suggest that the neurons with low-pass MTF should exhibit less or no threshold adaptation at the time scale required. Platkiewicz and Brette (2011) showed that the dynamics of sodium and potassium channels may or may not induce threshold adaptation, depending on their biophysical parameter values. Even if threshold adaptation is present, the corresponding time constants are not always within the range required to induce high-pass filtering.

Less is known about envelope coding in the avian auditory system. It was shown that barn owl IC neurons are sensitive to
amplitude modulation (Keller and Takahashi 2000), and that the MTF can be band pass downstream of CN (Nagel and Doupe 2006; Woolley and Casseday 2005). Band-pass MTFs are already prominent in NA, but are absent in NM (Steinberg and Peña 2011). While it is accepted that the pathway via NM conveys information used for sound localization through interaural time processing (Carr and Konishi 1990; Fukui et al. 2006; Köppl 1997; Oertel 1999; Sullivan and Konishi 1984), the role of NA is less clear (MacLeod and Carr 2007). Indeed, whereas the NM population is homogenous and corresponds closely to mammalian bushy cells, NA physiology is heterogeneous (Köppl and Carr 2003) with responses similar to the rest of the mammalian CN. Beside its role in coding sound level (Sullivan and Konishi 1984; Takahashi et al. 1984), we show here that NA can also encode the envelope at the output of the cochlear channels more efficiently than its input. The fact that the filtering depends on intensity suggests sound identification may depend on location. The gain is larger for sounds in the front due to the head-related transfer function, which could enhance the envelope encoding ability at this location (Steinberg et al. 2013). Our results thus support the hypothesis of an early segregation of timing and envelope pathways (Steinberg and Peña 2011).

ACKNOWLEDGMENTS

The authors thank Romain Brette for fruitful discussions and for comments on the manuscript, and Brian Fischer for help with the GLM implementation.

GRANTS

This work was supported by a European Community Marie Curie fellowship (PIOF-GA-2011-300753) for B. Fontaine; by the Deutsche Forschungsgemeinschaft (SFB/TRR31 “Active Hearing” to C. Köppl; by National Institute of Deafness and Other Communications Disorders Grant DC-007690 to J. L. Peña.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: B.F. and J.L.P. conception and design of research; B.F., K.M.M., S.T.L., C.K., and J.L.P. performed experiments; B.F. analyzed data; B.F. interpreted results of experiments; B.F. prepared figures; B.F. drafted manuscript; B.F., K.M.M., C.K., and J.L.P. edited and revised manuscript; B.F., K.M.M., S.T.L., J.L.S., C.K., and J.L.P. approved final version of manuscript.

REFERENCES


